SURPH 2.1

SURvival Under Proportional Hazards

Technical Manual
2003

DOE/BP-00012494-1

July 2003

This Document should be cited as follows:

Lady, James, Peter Westhagen, John Skalski, ''SURPH 2.1'', Project No. 1989-10700, 71 electronic pages, (BPA Report DOE/BP-00012494-1)

Bonneville Power Administration
P.O. Box 3621
Portland, Oregon 97208

This report was funded by the Bonneville Power Administration (BPA), U.S. Department of Energy, as part of BPA's program to protect, mitigate, and enhance fish and wildlife affected by the development and operation of hydroelectric facilities on the Columbia River and its tributaries. The views in this report are the author's and do not necessarily represent the views of BPA.

SURPH 2.1

SURvival under Proportional Hazards

Developed by
James Lady, Peter Westhagen, and John R. Skalski
School of Aquatic and Fishery Sciences
University of Washington
1325 4th Avenue, Suite 1820
Seattle, WA 98101-2509

Prepared for
U.S. Department of Energy
Bonneville Power Administration
Division of Fish and Wildlife
P.O. Box 3621
Portland, OR. 97208-3621
Project No. 8910700
Contract No. DE-B179-90BP02341

February 2002

About This Manual

This document is primarily a description of the user interface for SURPH2.1. It is not a description of the statistical theory and calculations behind SURPH. The best source for that information is the SURPH.1 manual. To request a copy of the SURPH.1 manual sent to you free of charge, please send the request, along with your mailing address, to surph@cbr.washington.edu.

Important! This document describes the user interface to SURPH 2.1. The format of the input data file has been changed since SURPH 1.0 and SURPH 2.0 to make it simpler and more intuitive. A conversion program to convert a data file from previous versions of SURPH to the input data file format for SURPH 2.1 is provided (Section 1.3.1).

This project is funded by the Bonneville Power Administration, US Department of Energy, under Contract No. 004126, Project No. 89-107-00, as part of the BPA's program to protect, mitigate, and enhance fish and wildlife affected by the development and operation of hydroelectric facilities on the Columbia River and its tributaries.

Table of Contents

About This Manual i
1.0 Getting Started 1
1.1 SURPH 2.1 - Startup 1
1.2 Select Data Type 1
1.3 SURPH 2.1 - Load Input Data 3
1.3.1 SURPH 2.1 - Data conversion utility 3
2.0 Basic Concepts 4
2.1 SURPH Options 4
2.2 SURPH Probabilities 4
2.3 Model Parameters 5
2.4 Covariates 5
2.4.1 Group covariates 5
2.4.2 Individual covariates 6
2.5 Link Functions 6
2.5.1 Link function details 7
2.5.2 Identity link 7
2.5.3 Hazard link 8
2.5.4 Logit link 8
2.6 SURPH 2.1 - Model Definition 8
2.7 SURPH 2.1 - Model 9
2.8 SURPH Dialog Window 9
2.8.1 Model definition list 10
2.8.2 Master model list 10
2.8.3 Testable model list 10
2.8.4 Create model definition button 10
2.8.5 Create model(s) button 10
2.8.6 Description of column labels 11
2.9 SURPH Reports 11
2.10 SURPH Graphs 12
3.0 Data Analysis Tools 13
3.1 Release-Recapture 13
3.1.1 Cormack-Jolly-Seber estimates 13
3.1.2 Profile likelihood confidence intervals 14
3.1.3 CJS covariance matrix 15
3.1.4 Burnham et al. (1987) tests 16
3.1.5 Manly-Parr estimates 17
3.1.6 Reduced M-arrays 18
3.1.7 Data analysis - individual covariates 19
3.1.7.1 Individual covariate - Histogram 19
3.1.7.2 Individual covariate – Cumulative distribution function 20

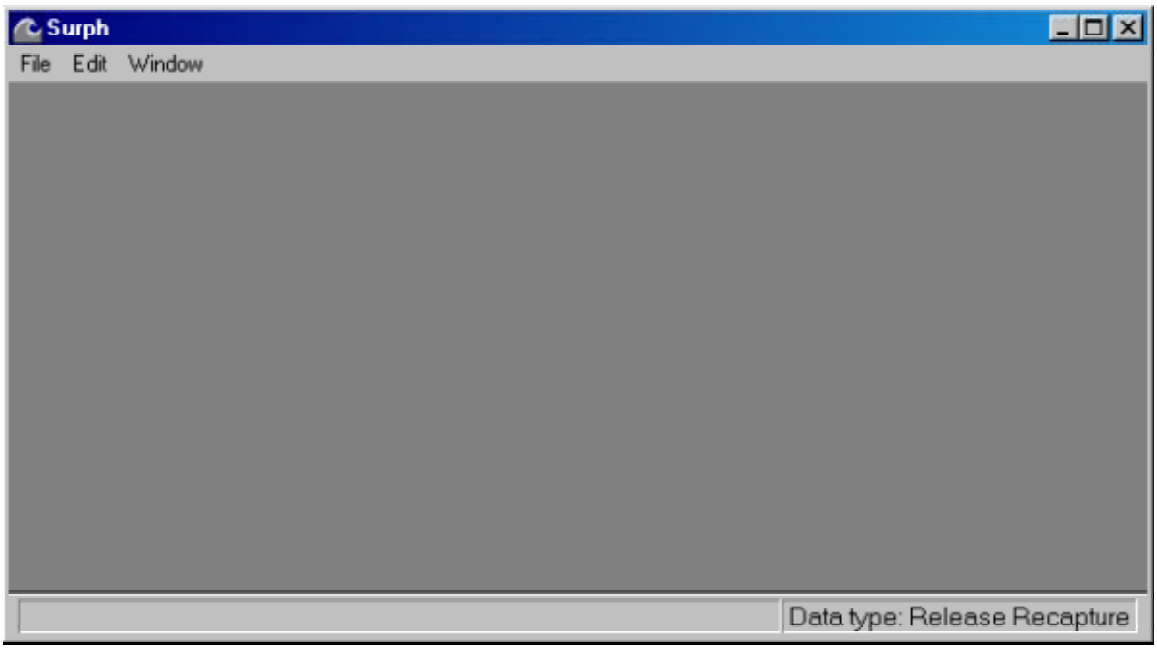
3.2 Known-Fate 22
3.2.1 Binomial estimates 22
3.2.2 Profile likelihood confidence intervals 23
3.2.3 Sampling summary 24
3.2.4 Data analysis - Individual covariates 25
3.2.4.1 Individual covariate - Histogram 25
3.2.4.2 Individual covariate – Cumulative distribution function 26
4.0 Modeling 28
4.1 Link Functions 29
4.2 Parameter Definitions 30
4.2.1 Main effects parameter definitions 30
4.2.2 Group covariate parameter definitions 34
4.2.3 Individual covariate parameter definitions 35
4.2.4 Define model - Release-recapture notes 36
4.3 Calculate Model Variance 37
4.4 Model Output 37
4.4.1 SURPH parameters report 38
4.4.2 SURPH probabilities report 39
4.5 Hypothesis Testing 39
4.5.1 Hypothesis testing - Likelihood ratio test 40
4.5.2 Hypothesis testing - Akaike information criterion 41
4.5.3 Hypothesis testing - Analysis of deviance 41
4.5.3.1 Analysis of deviance - Initialization 42
4.5.3.2 Analysis of deviance - Components 44
4.6 Model Diagnostics 46
4.6.1 Model diagnostics - Group covariate survival curves 47
4.6.2 Model diagnostics - Individual covariate survival curves 48
4.6.2.1 Histogram 49
4.6.2.2 Nonparametric estimate 50
4.6.2.3 Relative Risk 51
4.6.2.4 Relative Risk with Histogram 52
4.6.3 Residuals vs. capture histories 54
4.6.4 Scatter option 56
4.6.5 Normal QQ plot 57
4.7 File Formats 58
4.7.1 Input file 58
4.7.2 Data description 58
4.7.3 Number of populations 58
4.7.4 Number of periods 58
4.7.5 Group covariate definitions 59
4.7.6 Individual covariate definitions 59
4.7.7 Population names 59
4.7.8 Number tagged 59
4.7.9 Tag ID flag 60
4.7.10 Capture history line 60
4.7.11 Example: SURPH input data file 60

4.8 File Operations 62
4.8.1 SURPH 2.1 - Load input data 62
4.8.2 Save and load model definitions 62
4.8.3 Save and load models 63
5.0 References 64

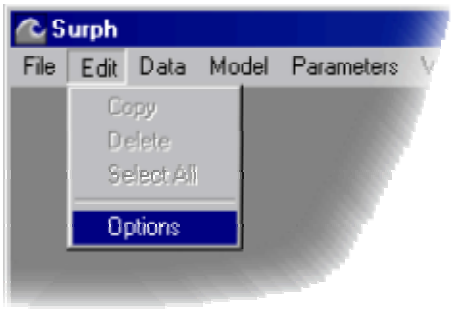
1.0 Getting Started

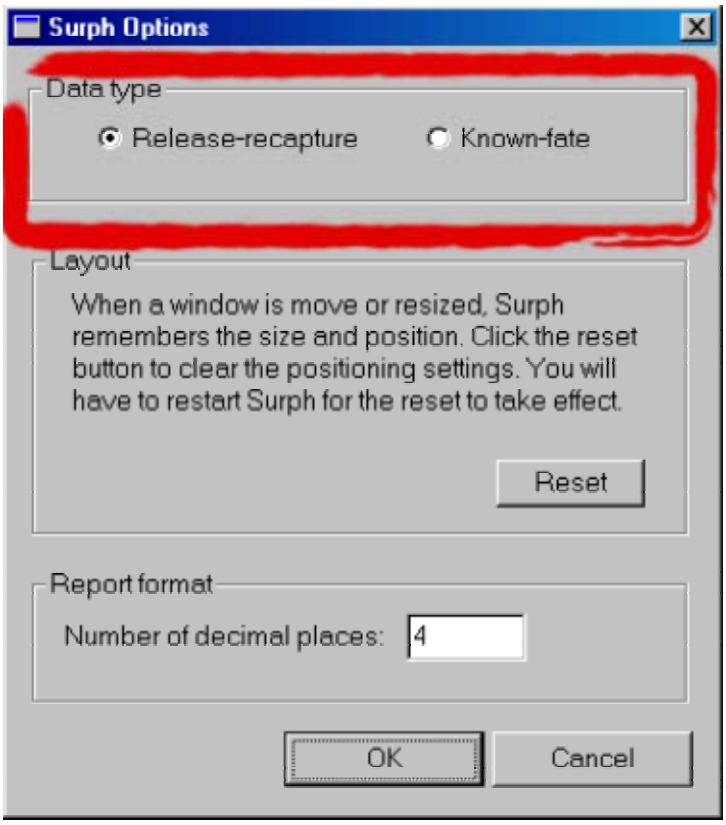
1.1 SURPH 2.1 - Startup

The SURPH 2.1 program at startup appears as shown below. The main SURPH window consists of a context-sensitive menu and a workspace that holds all open windows during a SURPH 2.1 session.

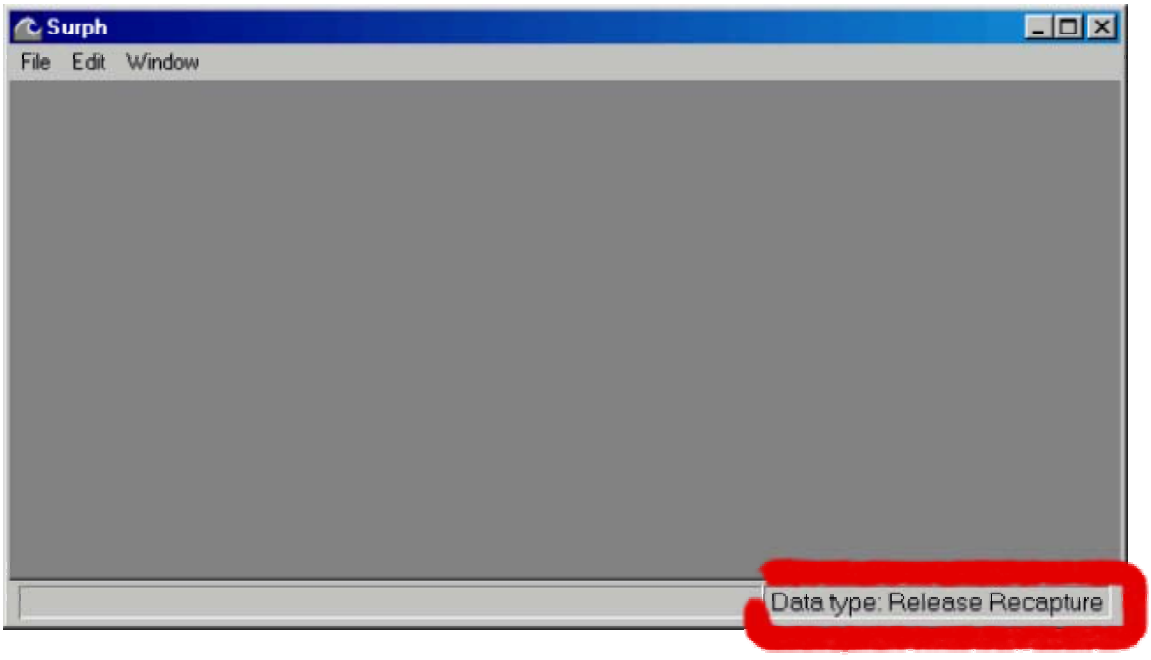


1.2 Select Data Type

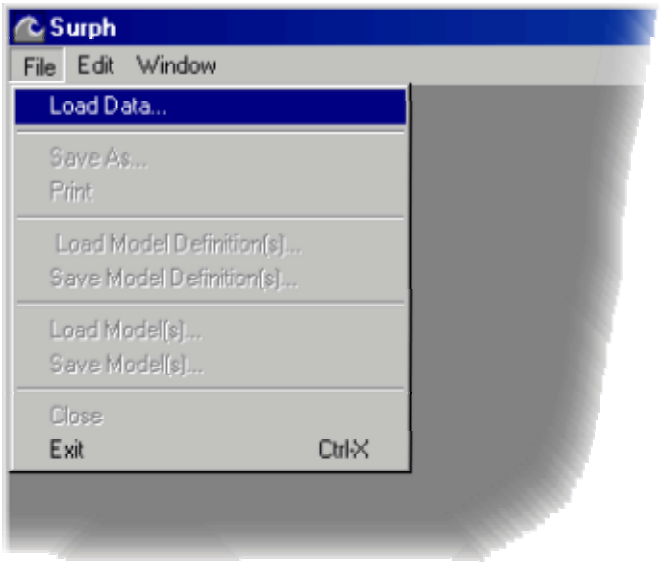




The currently selected data type appears in the lower right-hand corner of the main SURPH window as shown below. Once the data type has been selected and SURPH 2.1 has exited normally, the selected data type will be the default at startup.



1.3 SURPH 2.1 - Load Input Data



Once the data have been successfully loaded, the SURPH Dialog Window for the selected data type will appear.

Note: The input data file format is different for SURPH 2.1 than for previous versions of SURPH 1.0 and SURPH 2.0. Please see Section 1.3.1 for detailed instructions on data input formatting. A conversion utility is provided to convert older input data files into the new format.

1.3.1 SURPH 2.1 - Data conversion utility

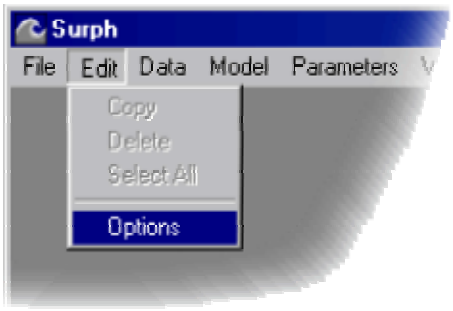
The stand-alone conversion routine "dataConv" will convert data files created for SURPH.1 or SURPH.2 into the appropriate format for SURPH 2.1. The task must be run from the command prompt in Windows 2000 or the MS-DOS prompt in Windows NT/98. The syntax is as follows:

- `dataConv oldDataFile newDataFile`

where oldDataFile is the input data file for SURPH.1 or SURPH.2, and newDataFile is the new data file to be created for SURPH 2.1.

2.0 Basic Concepts

2.1 SURPH Options



The SURPH Options dialog allows the user to select options that affect the overall behavior of SURPH, as follows:

- The **Data Type selection** determines the type of data that will be analyzed by SURPH: Known-Fate or Release Recapture.
- The **Reset** option allows the user to reset the window positions to their default position upon opening.
- The **Report format - Number of decimal places** determines the number of decimal places on all values for SURPH reports.

2.2 SURPH Probabilities

The **SURPH probabilities** are the probabilities of interest in a SURPH model. Specifically, they are:

- The survival probabilities for known-fate studies.
- The survival probabilities, as well as the detection probabilities, and the final product probabilities for release-recapture studies.

The SURPH probabilities are modeled as a function of the model parameters and the covariate values. The functional relationship of the parameters and covariates to the resulting SURPH probabilities is defined by the link function.

2.3 Model Parameters

SURPH 2.1 models the SURPH probabilities as a function of the **model parameters**. The model parameters are:

- The **baseline probability** defines the given probability for the first population and period.
- The **main effects** describe the difference of the given SURPH probability from the baseline probability.
- The **group covariate coefficients** describe the relationship between the group covariates and the SURPH probability.
- The **individual covariate coefficients** describe the relationship between the individual covariates and the SURPH probability.

2.4 Covariates

SURPH probabilities can be modeled as a function of two types of covariates:

1. Group Covariates.
2. Individual Covariates.

2.4.1 Group covariates

Group covariates are factors that affect one or more of the populations as a whole. Examples are ambient temperature, sex, habitat type, or treatment condition. There are two varieties of group covariates:

a. **Time-invariant** group covariates that are the same from one sampling occasion to the next. The user defines the value of a time-invariant group covariate for each population. An example might be the sex of the individuals in the populations.

b. **Time-variant** group covariates are allowed to change from one sampling occasion to the next. The user defines the value of a time-variant group covariate for each period of each population. An example might be the average temperature for each period.

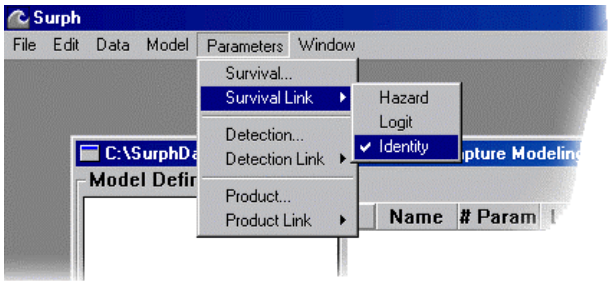
2.4.2 Individual covariates

The Individual covariates apply to each individual. Examples include body length and body weight. The user specifies the value of an individual covariate for each individual.

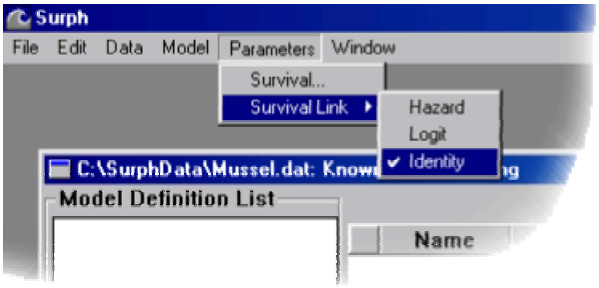
The nature of the relationship between the covariates and the resulting SURPH probabilities is defined by the link function.

2.5 Link Functions

Release-Recapture



Known-Fate



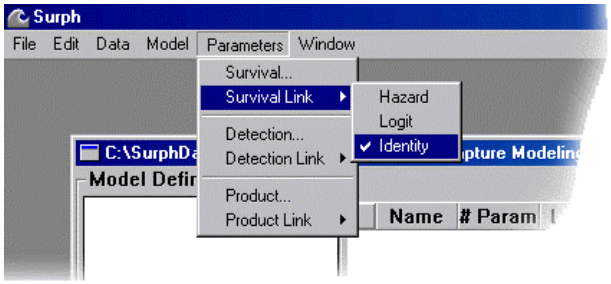
The functional relationship of the model parameters to the SURPH probabilities is defined by the link function. SURPH 2.1 defines three link functions.

- The **identity link** assumes an additive effect on survival probability. The identity link is the default link for all SURPH probabilities.
- The **hazard link** assumes a multiplicative effect on hazard rate, an exponential effect on survival probability.
- The **logit link** assumes an additive effect on log odds ratio.

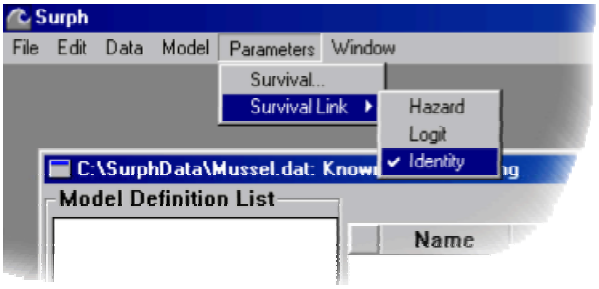
Both the hazard link and the logit link constrain all probabilities to be between 0 and 1; the identity link allows a probability to be greater than 1.

2.5.1 Link function details

Release-Recapture



Known-Fate



2.5.2 Identity link

The identity link assumes an additive effect and places no constraints on the resulting SURPH probability. Survival is modeled as follows:

S_{ijk} = The survival probability for population *i*, period/occasion *j*, individual *k*.

S_0 = The baseline survival probability parameter.

π_{ij} = The main effect parameter for population i, period/occasion j.

γ_j = The vector of group covariate coefficients for period/occasion j.

x_{ij} = The vector of group covariates for population i, period/occasion j.

δ_{ij} = The vector of individual covariate coefficients for population i, period j.

y_{ik} = The vector of individual covariates for population *i*, individual *k*.

2.5.3 Hazard link

The hazard link assumes a multiplicative effect on hazard rate and an exponential effect on the SURPH probability. The hazard link constrains the resulting SURPH probability to be between 0 and 1.

$$S_{ijk} = \text{logit}\left(S_0\right)^{\exp\left(\pi_{ij}+\gamma_j x_{ij}+\delta_{ij} y_{ik}\right)}$$

2.5.4 Logit link

The logit link assumes an additive effect on the log odds ratio. The logit link, like the hazard link, constrains the resulting SURPH probability to be between 0 and 1

$$S_{ijk} = \frac{\exp\left(S_0 + \pi_{ij} + \gamma_j x_{ij} + \delta_{ij} y_{ik}\right)}{1+\exp\left(S_0 + \pi_{ij} + \gamma_j x_{ij} + \delta_{ij} y_{ik}\right)}$$

2.6 SURPH 2.1 - Model Definition

A **Model Definition** encapsulates all the settings that go into defining a model, as follows:

1. The name of the model.
2. The settings for the main effects, group covariates, and individual covariates for:
    - Survival-related parameters (known-fate and release-recapture data).
    - Detection-related parameters (release-recapture data only).
    - Final product-related parameters (release-recapture data only).

The use of model definitions allows more than one model to be defined in advance, and the actual model creation (i.e., parameter estimation) to be performed as a separate operation. For some models, especially those with individual covariate parameters, a significant amount of time is required to estimate the parameters; hence, there is an advantage to creating several model definitions in advance and estimating the parameters for all models in one operation.

2.7 SURPH 2.1 - Model

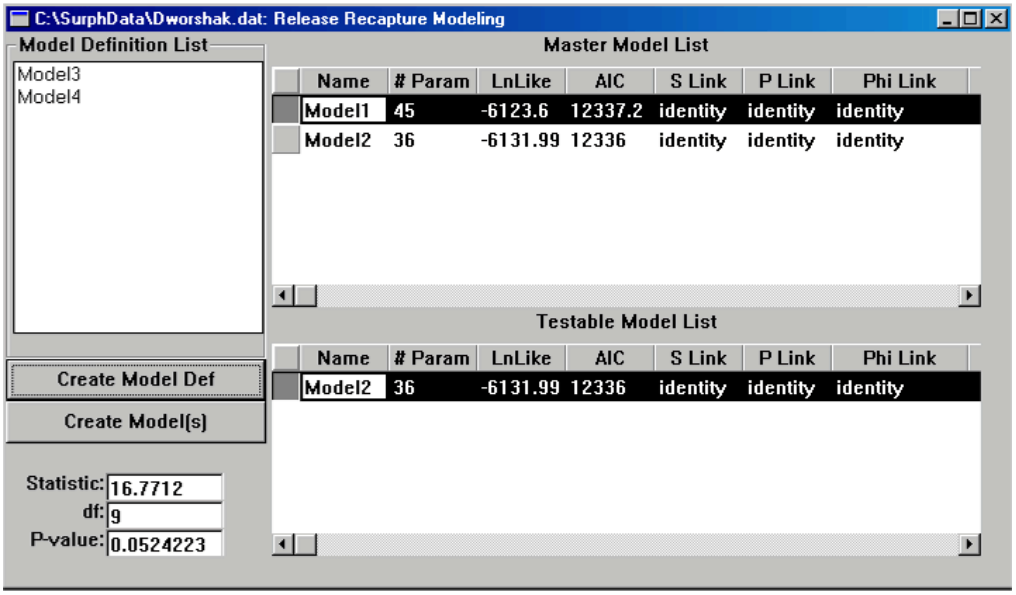
A **Model** is created when the parameters, as defined by the model definition, are estimated. A Model thus contains all the information contained in a model definition, plus the parameter estimates. Optionally, the variances and covariances of the parameter estimates can also be computed upon request.

2.8 SURPH Dialog Window

The SURPH Dialog Window for release-recapture data is shown below (the SURPH dialog window for known-fate data differs only in the column labels in the main model list and the testable model list). Once input data has been loaded into SURPH 2.1, the SURPH dialog window is the base window for all data analysis and modeling. The main components are:

- The Model Definition List.
- The Master Model List
- The Testable Model List
- The Likelihood Ratio Test (LRT) area
- The Create Model Definition button
- The Create Model(s) button

Description of column labels



2.8.1 Model definition list

The Model Definition List contains the name of all model definitions that have not yet been used to create models. In the above example, pressing the "Create Model" button will cause the creation of models "Model3" and "Model4" from their respective Model Definitions. The models will then appear in the Master Model List and their definitions will be removed from the Model Definition List. Model Definitions can be exported and imported at a later SURPH 2.1 session.

2.8.2 Master model list

The Master Model List contains all the models that have been created in the current session of SURPH 2.1. Models can be exported and imported at a later SURPH 2.1 session.

2.8.3 Testable model list

When a model is selected in the Master Model List, any nested models of the selected model that are in the Master Model List appear in the Testable Model List as well. If the user selects one of these models in the Testable Model List, a Likelihood Ratio Test (LRT) is performed, and the results appear in the Likelihood Ratio Test area. In the above example, "Model2" is selected in the Testable Model list and "Model1" is selected in the Main Model List. The results of the LRT are in the Likelihood Ratio Test area.

2.8.4 Create model definition button

When the Create Model Definition button is pressed, the user will be prompted to enter a name for the Model Definition about to be created from the current parameter settings. A unique name must be entered. A Model Definition cannot be created with the same settings as an already defined Model Definition or Model.

2.8.5 Create model(s) button

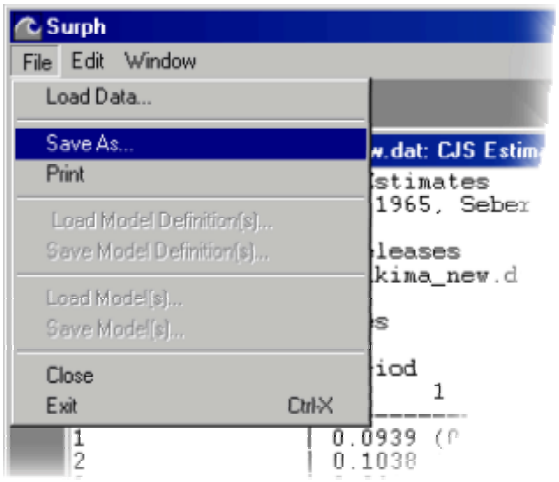
The Create Model(s) button creates models from all of the Model Definitions in the Model Definition List. The names of these models then appear in the Master Model List, and their names are removed from the Model Definition List.

2.8.6 Description of column labels

Name	The model name
# Param	The number of parameters in the model
LnLike	The log-likelihood
AIC	The Akaike Information Criterion
S Link	The link function for the survival probabilities
P Link	(Release-recapture data only) The link function for the detection probabilities
Phi Link	(Release-recapture data only) The link function for the final product probabilities

2.9 SURPH Reports

One way in which SURPH displays results is via text-based reports. In all reports, the user has the following options:



Save As . . .	Save the report to a text file. The resulting file can be edited and/or printed with any standard text editor.
Print	Print the report.

The precision of the report values can be set through the SURPH Options dialog.

2.10 SURPH Graphs

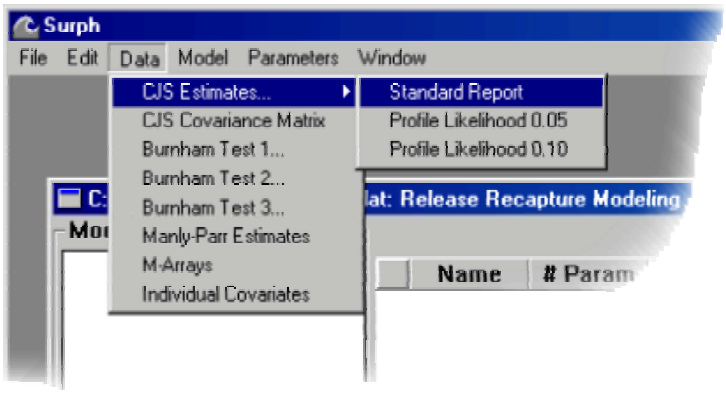
In several cases, SURPH 2.1 displays results in a graphical form. When a graph is selected, an "Option" menu appears on the menu bar with some or all of the following menu selections.

Model Selection	Description
Grid Line	Controls the density of grid lines on the graph.
Zoom	Allows the user to zoom in to the graphical data to varying degrees.
Move Origin	Allows the user the move the center of the graph frame to a different point on the graph.

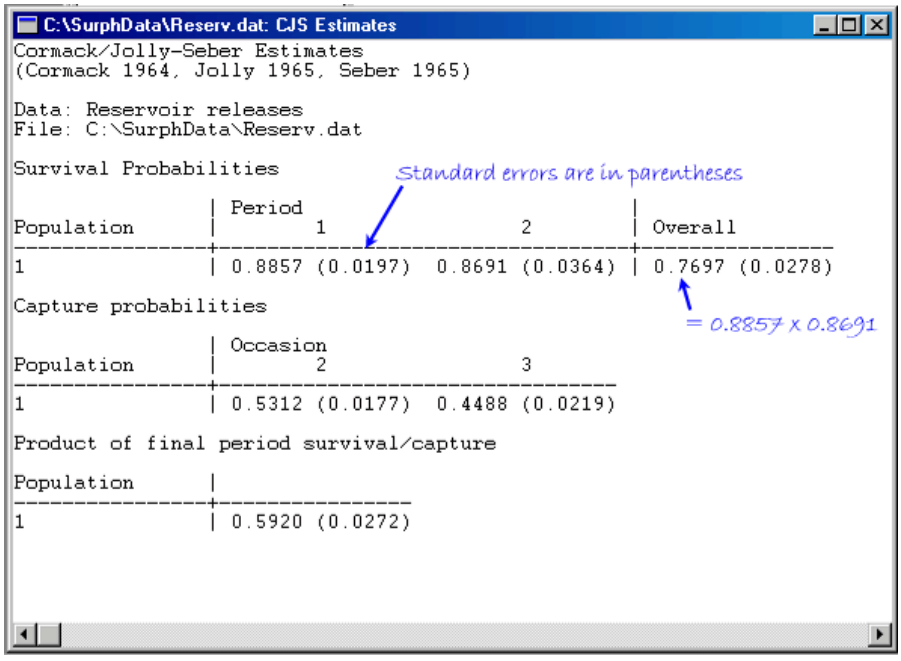
3.0 Data Analysis Tools

3.1 Release-Recapture

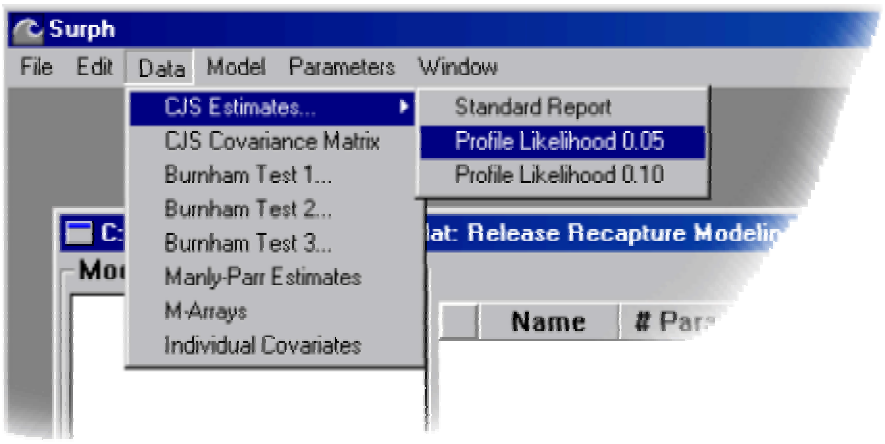
3.1.1 Cormack-Jolly-Seber estimates



The Cormack-Jolly-Seber (CJS) estimates report, as shown below, provides estimates and standard errors for the survival and detection probabilities, and the product of the final survival and detection probabilities. The standard error of all estimates are reported in parentheses. The estimate of the overall survival probability—that is, the product of all the survival estimates—is reported as well. In the example below, there is one population with three periods.



3.1.2 Profile likelihood confidence intervals



If one of the profile likelihood options is selected for the CJS estimates, the report will include the estimates of the profile likelihood confidence intervals for all survival estimates. The alpha level for the confidence interval can be selected to be 0.05 or 0.10. The report below illustrates the CJS report with 95% profile likelihood confidence intervals for a study with three populations and two periods.

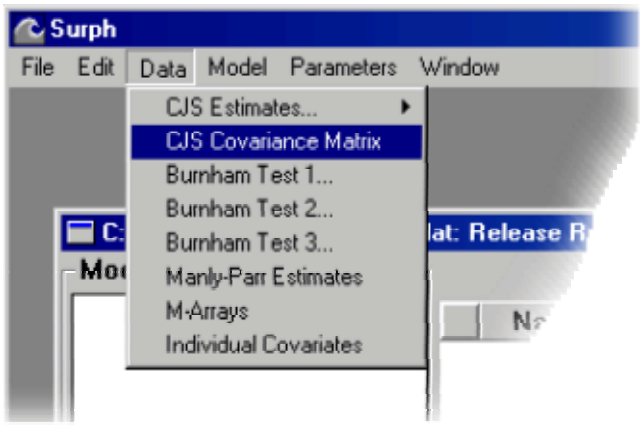
```
C:\SurphData\yakima.dat: CJS Estimates
Cormack/Jolly-Seber Estimates
(Cormack 1964, Jolly 1965, Seber 1965)
with profile likelihood 0.05 confidence intervals

Data: Yakima: Lake Releases
File: C:\SurphData\yakima.dat

Survival Probabilities

                | Period                                  |
Population      |                1                        | Overall
----------------+-----------------------------------------+-----------------------------------
1               | 0.0939 (0.0115) [0.0743 0.1208]         | 0.0939 (0.0115)
2               | 0.1038 (0.0112) [0.0840 0.1292]         | 0.1038 (0.0112)
3               | 0.0619 (0.0086) [0.0472 0.0820]         | 0.0619 (0.0086)

Capture probabilities

                | Occasion
Population      |                2
----------------+-----------------------------------
1               | 0.5741 (0.0673)
2               | 0.5405 (0.0579)
3               | 0.5556 (0.0741)

Product of final period survival/capture

Population      |
----------------+-----------------------------------
1               | 0.3827 (0.0540)
2               | 0.4878 (0.0552)
3               | 0.4902 (0.0700)
```

alpha level

lower bound of 95% confidence interval

upper bound

3.1.3 CJS covariance matrix

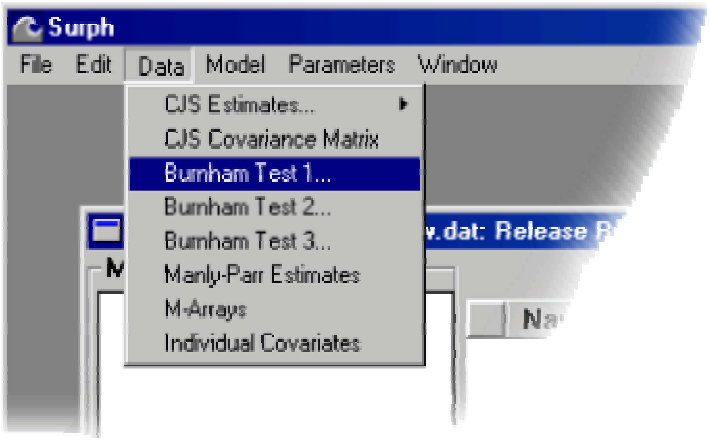


The CJS Covariance Matrix report displays the covariance matrix for all of the Cormack-Jolly-Seber parameters, as shown in the example below. The values are printed in exponential notations with full precision for use by other analysis programs; the number of decimal places in the Report Format options has no effect on this report.

C:\SurphData\yakima_new.dat: Cjs Covariance Report

```
Cormack/Jolly-Seber covariance matrix
(Cormack 1964, Jolly 1965, Seber 1965)

Data: Yakima: Lake Releases
File: C:\SurphData\yakima_new.dat

        S1                  P2

------------------
Population: 1


S1         1.315106e-004      -4.573750e-004
P2        -4.573750e-004       4.528019e-003

------------------
Population: 2


S1         1.271202e-004      -3.302070e-004
P2        -3.302070e-004       3.356168e-003

------------------
Population: 3


S1         7.378746e-005      -3.114705e-004
P2        -3.114705e-004       5.486968e-003
```

3.1.4 Burnham et al. (1987) tests



Burnham et al. (1987) gives three types of statistical tests for assessing model assumptions for the release-recapture protocol.

1. **TEST 1**: Tests for overall treatment effects.

   H_0: All survival and detection parameters are the same across treatment groups.

   H_A: At least some parameters differ between or among groups.

2. **TEST 2**: Tests for goodness-of-fit of the model to the data.

   H_0: The survival and detection parameters are specific to sampling occasions within each group.

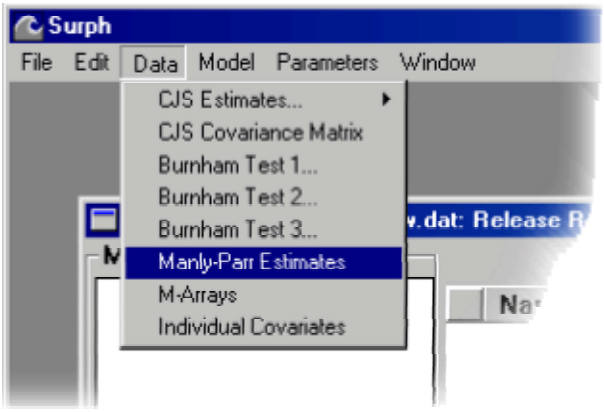
   H_A: The model does not fit the data; there may be tagging effects or differential behavior.

3. **TEST 3**: Tests for parameters that are specific to individual capture histories.

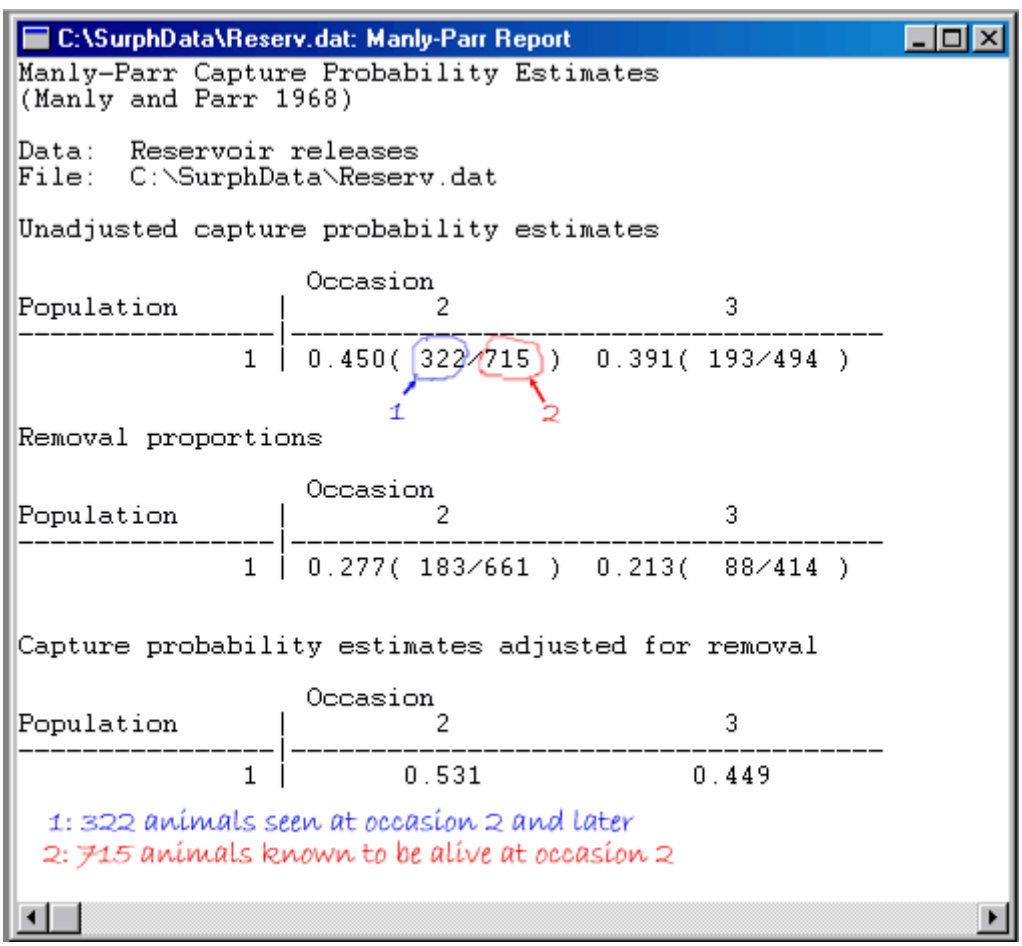
   H_0: The survival and detection parameters do not depend on the capture histories of fish released on any release occasion.

   H_A: Some of the parameters are dependent on the capture histories of fish in a given release.

3.1.5 Manly-Parr estimates



The Manly-Parr estimates (Manly and Parr 1968) provide a way of estimating detection probabilities independent of survival. The Manly-Parr report gives these estimates of detection probability as show below.

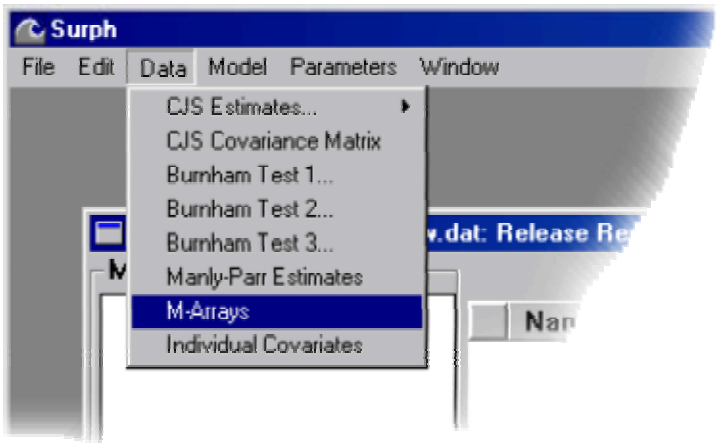


C:\SurphData\Reserv.dat: Manly-Parr Report

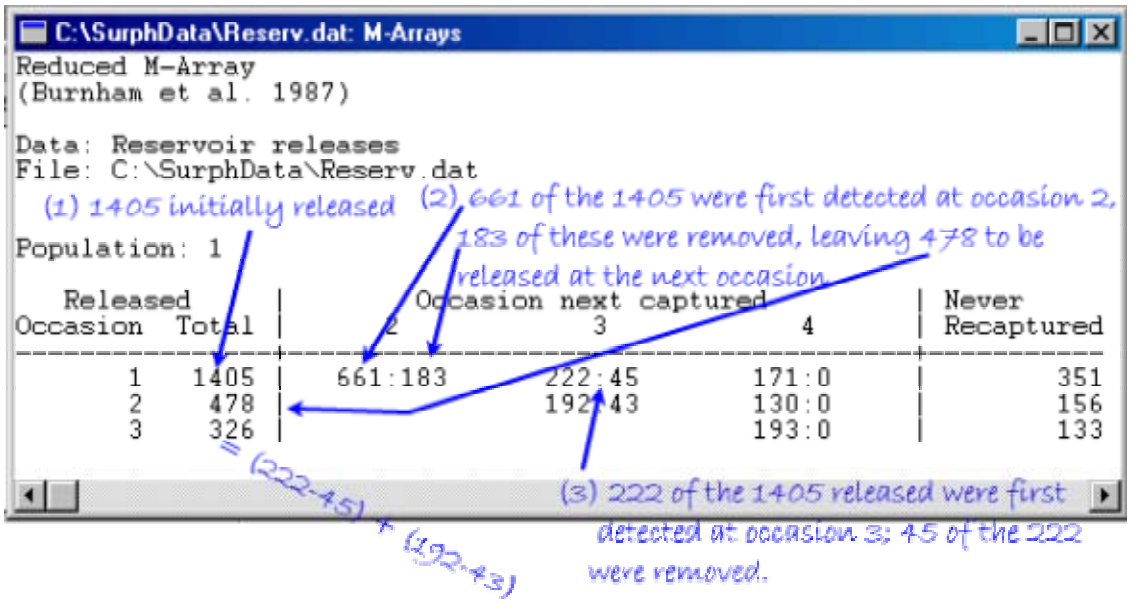
```
Manly-Parr Capture Probability Estimates
(Manly and Parr 1968)

Data:  Reservoir releases
File:  C:\SurphData\Reserv.dat

Unadjusted capture probability estimates

                  Occasion
Population       |         2                  3
----------------|-------------------------------------
              1 | 0.450( 322/715 )  0.391( 193/494 )
                        1        2

Removal proportions

                  Occasion
Population       |         2                  3
----------------|-------------------------------------
              1 | 0.277( 183/661 )  0.213(  88/414 )

Capture probability estimates adjusted for removal

                  Occasion
Population       |         2                  3
----------------|-------------------------------------
              1 |       0.531              0.449
```

1: 322 animals seen at occasion 2 and later
2: 715 animals known to be alive at occasion 2

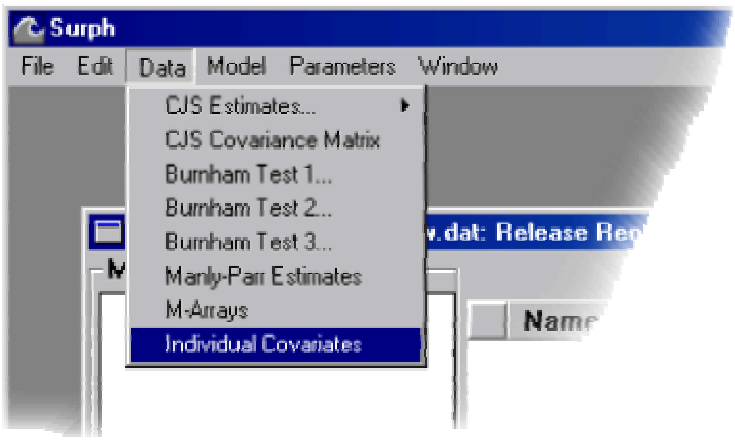
3.1.6 Reduced M-arrays



The "reduced M-arrays" of Burnham et al. (1987) display the numbers of animals released and subsequently recaptured. It also shows the number "removed" (removed or lost to handling mortality) at each occasion. There is an M-Array for each population in the study.



3.1.7 Data analysis - individual covariates

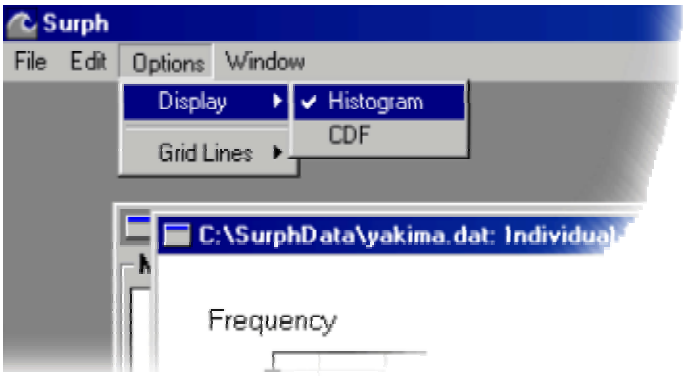


When the data contain covariate information for each individual in the study (individual covariates), SURPH 2.1 provides tools to analyze the covariate data.

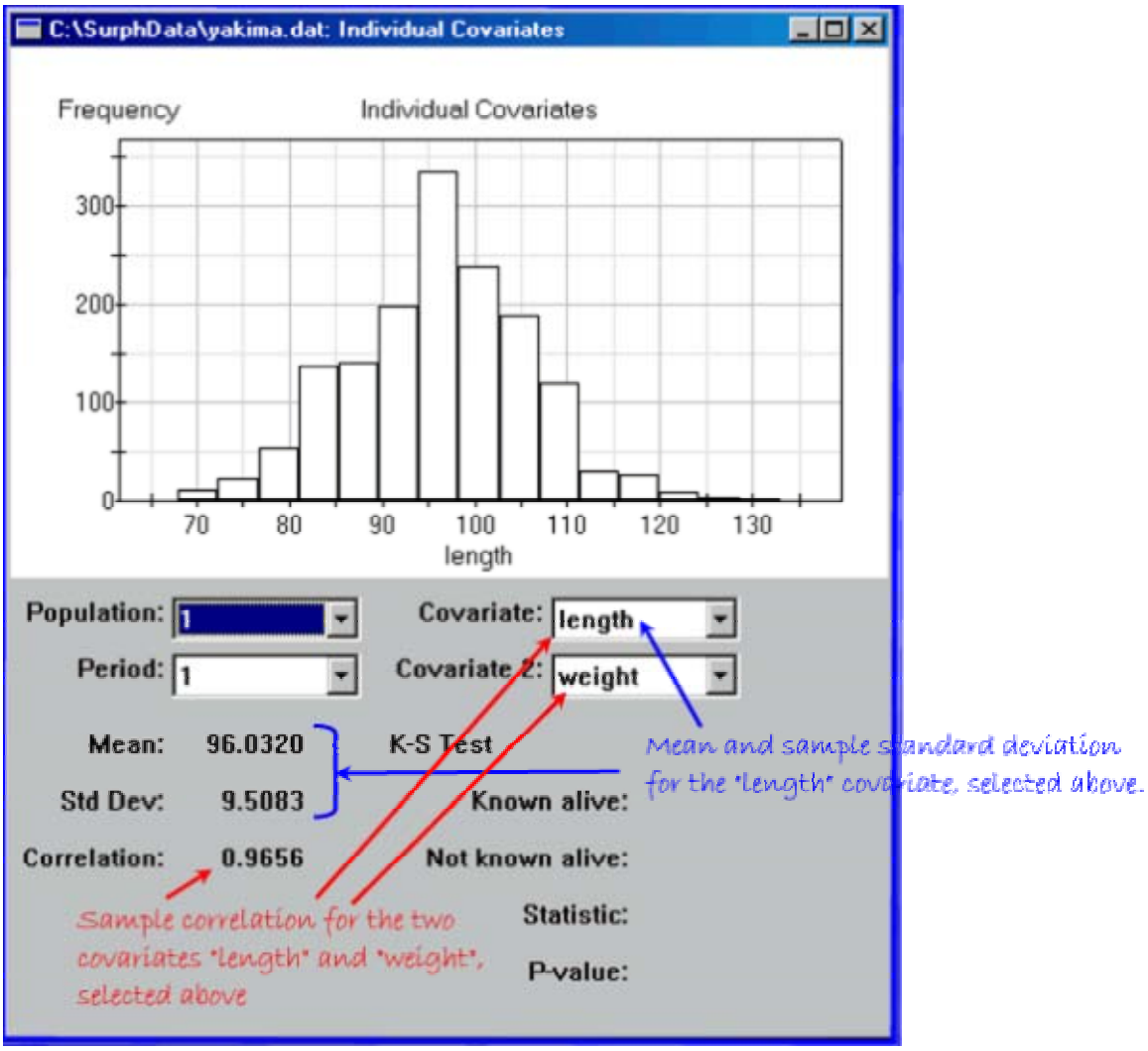
The user can select one of two ways of displaying the individual covariate data:

1. Histogram.
2. Cumulative distribution function.

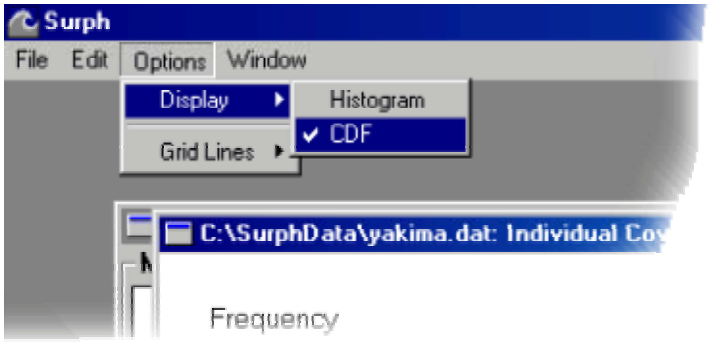
3.1.7.1 Individual covariate - Histogram



The histogram is the default display when the individual covariate option is selected from the data menu. In the menu shown above, the user can toggle between the histogram display and the cumulative distribution function display.



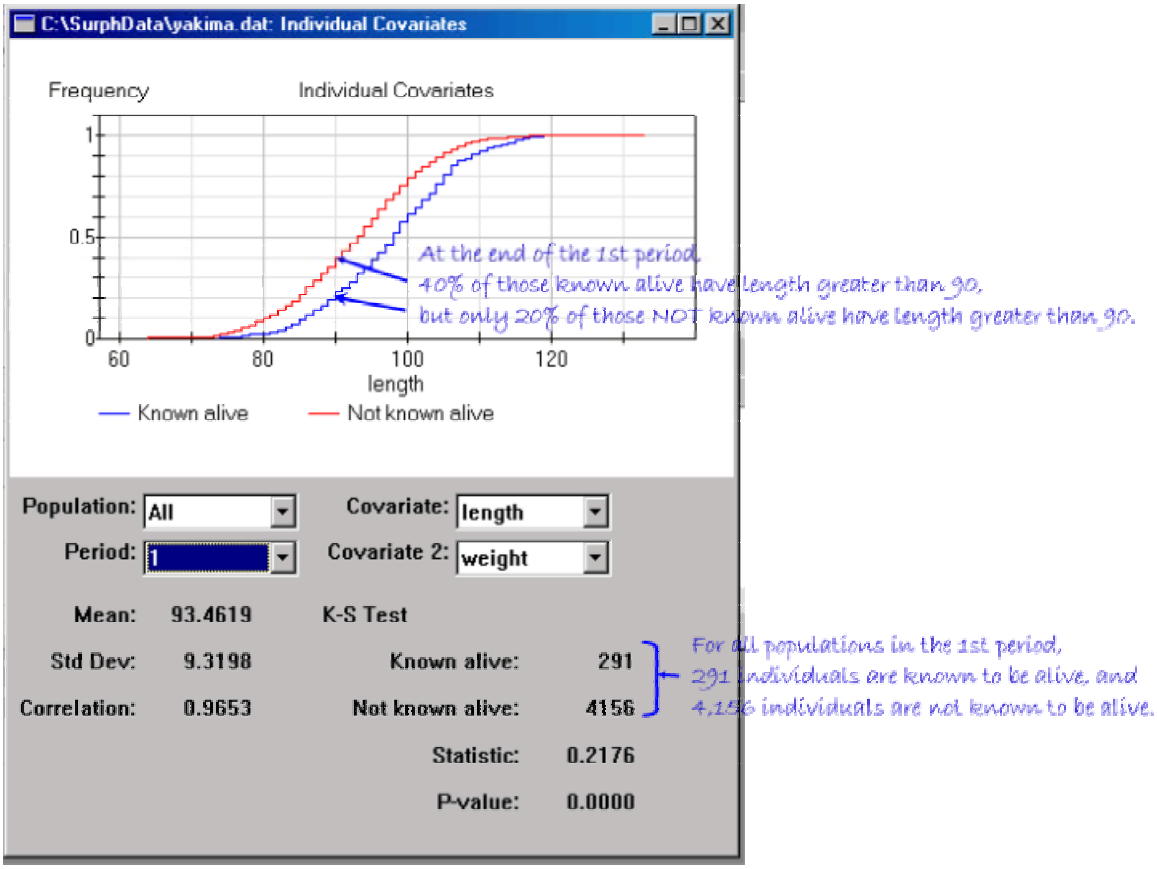
3.1.7.2 Individual covariate – Cumulative distribution function



The cumulative distribution function graph displays two cumulative distributions:

1. One for previously marked animals that were known to be alive at the end of the selected period (i.e., recaptured at the end of the period or later) in blue.

2. One for those not known to be alive (i.e., captured for the last time prior to the selected period) in red.



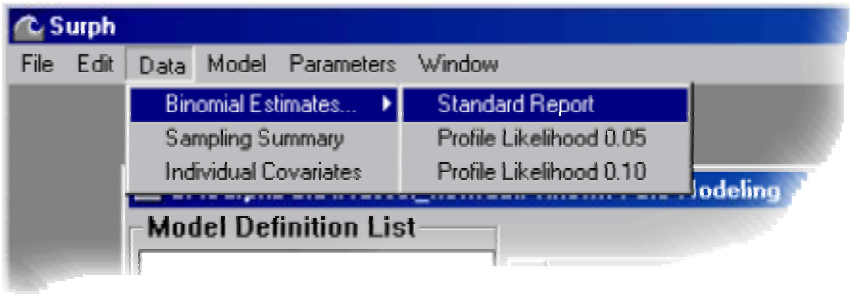
The Kolmogorov-Smirnov test, or K-S test, tests the following:

- H_0: The two samples are from the same distribution.
- H_A: The two samples are from different distributions.

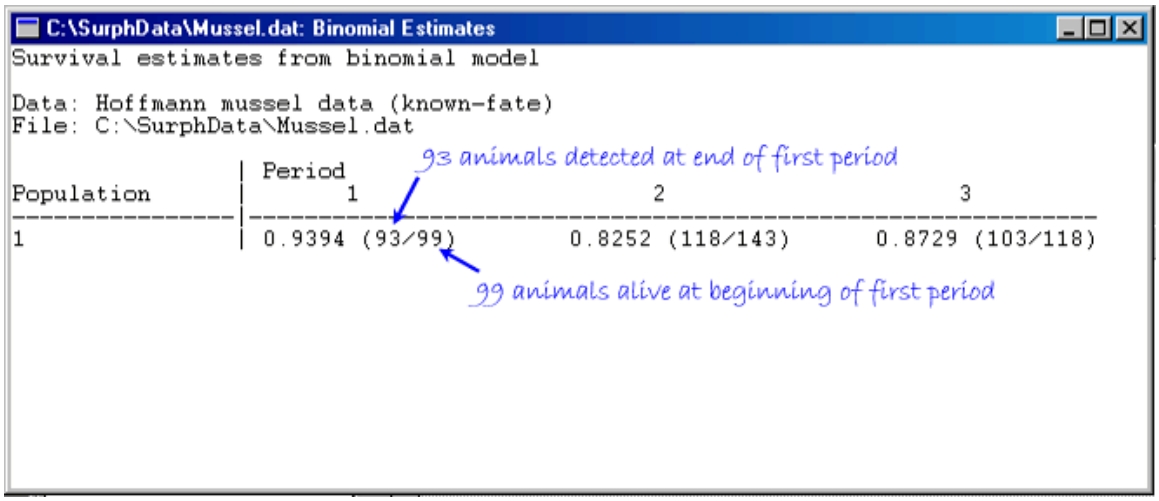
For the example above, the K-S test has a p-value of 0.000, indicating a significant difference in the two distributions. This may be due to some selective effect of length on survival, capture, or both .

3.2 Known-Fate

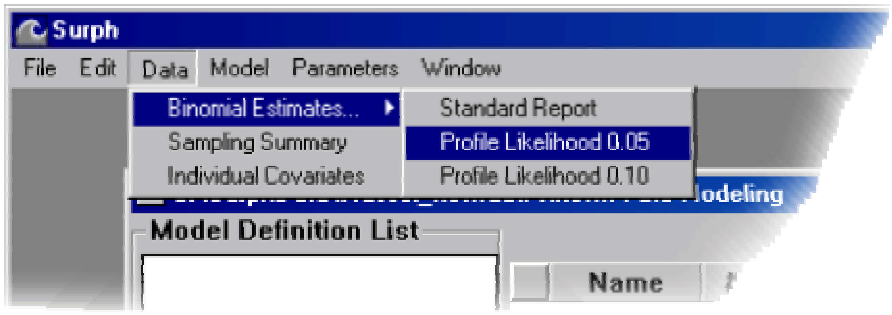
3.2.1 Binomial estimates



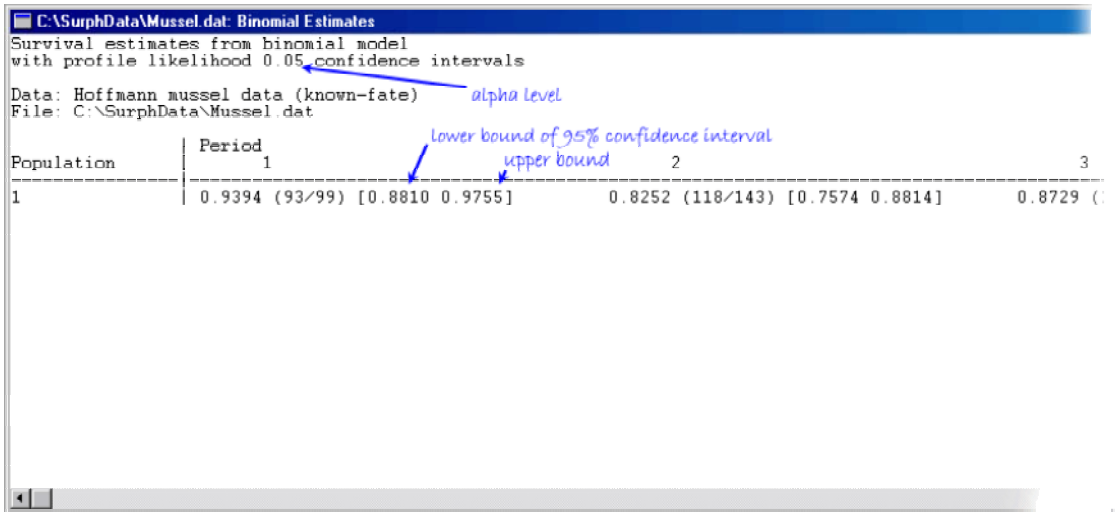
The binomial estimates report, shown below, provides survival estimates for each population and each period.



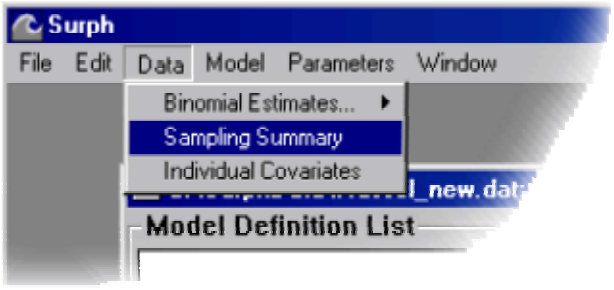
3.2.2 Profile likelihood confidence intervals



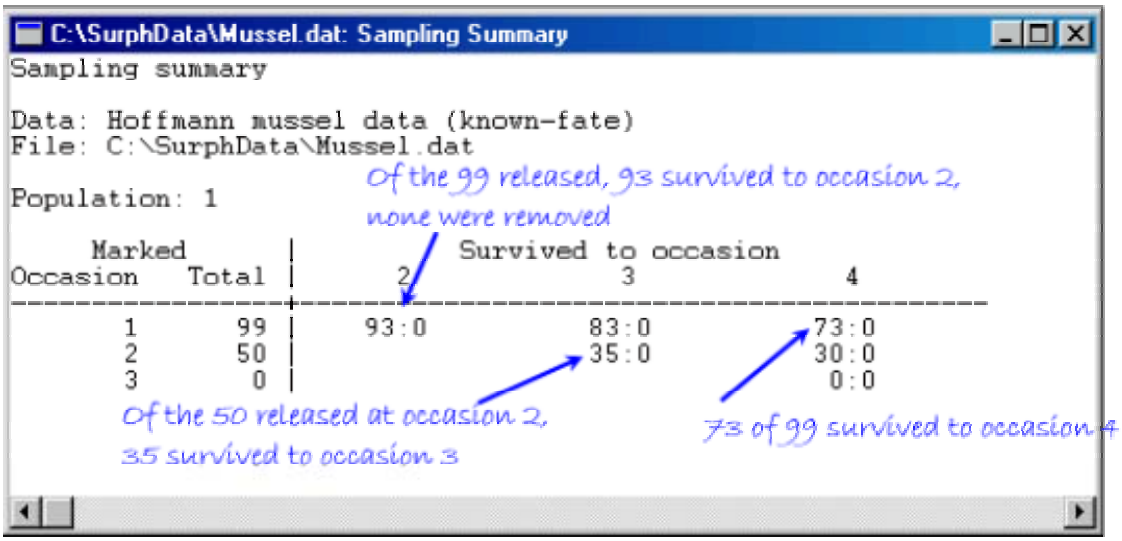
If one of the profile likelihood options is selected for the Binomial estimates, the report will include the estimates of the profile likelihood confidence intervals for all survival estimates. The alpha level for the confidence interval can be selected to be 0.05 or 0.10. The report below illustrates the Binomial report with 95% profile likelihood confidence intervals for a study with one population and three periods.



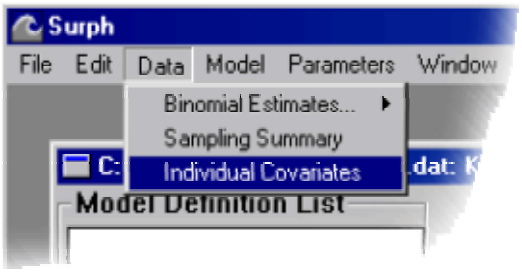
3.2.3 Sampling summary



The sampling summary shows the number of animals released at each occasion, and how many of each release survive to each subsequent occasion. It also shows the number "removed" (removed or lost to handling mortality) at each occasion. There is an sampling summary table for each population in the study.



3.2.4 Data analysis - Individual covariates

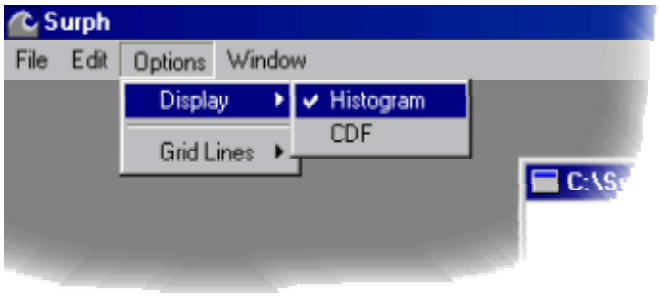


When the data contain covariate information for each individual in the study (individual covariates), SURPH 2.1 provides tools to analyze the covariate data.

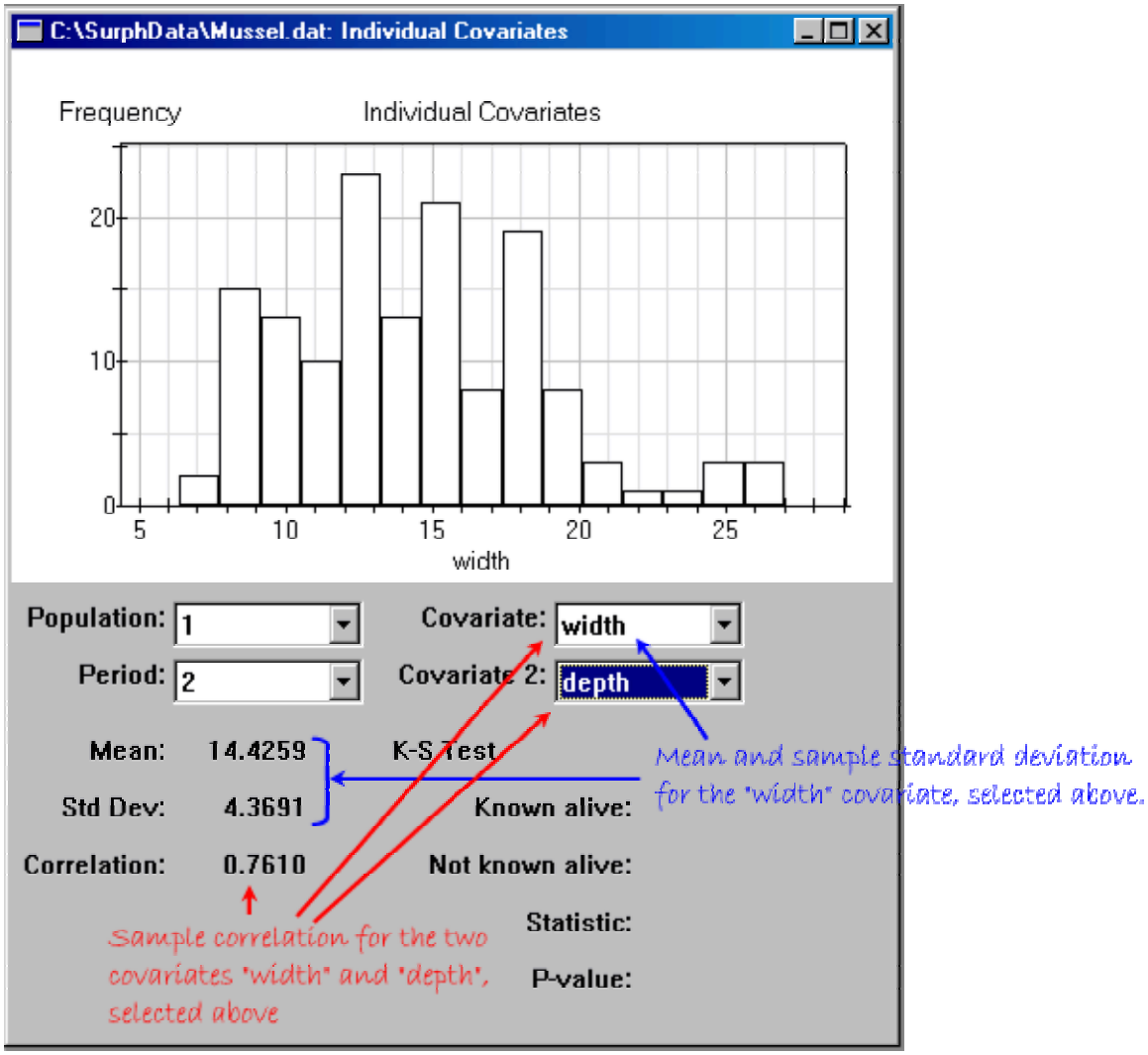
The user can select one of two ways of displaying the individual covariate data:

1. Histogram.
2. Cumulative distribution function.

3.2.4.1 Individual covariate - Histogram



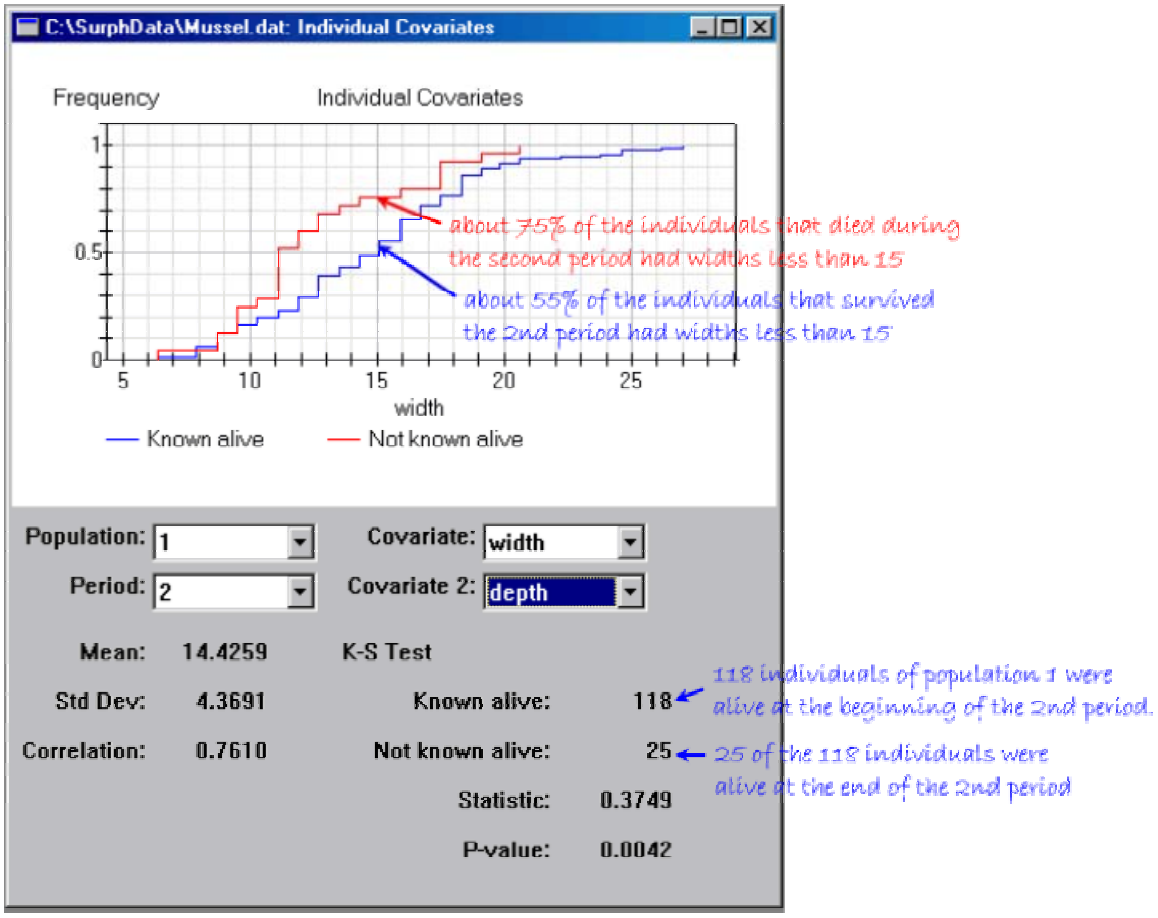
The histogram is the default display when the individual covariate option is selected from the data menu. In the menu shown above, the user can toggle between the histogram display and the cumulative distribution function display.



3.2.4.2 Individual covariate – Cumulative distribution function

The cumulative distribution function graph displays two cumulative distributions:

1. One for previously marked animals that were alive at the end of the selected period in blue.
2. One for those that died during the selected period in red.



For the example above, the K-S test has a p-value of 0.0042, indicating a significant difference in the two distributions. This may indicate some selective effect of length on survival .

4.0 Modeling

A central objective of SURPH is to characterize the survival probabilities (or, more generally, any SURPH probabilities), in terms of their relationship to:

- Each other across periods and/or populations via the main effects parameters.
- Measured covariates via the group covariate coefficients and the individual covariate coefficients.

The steps of modeling in SURPH are:

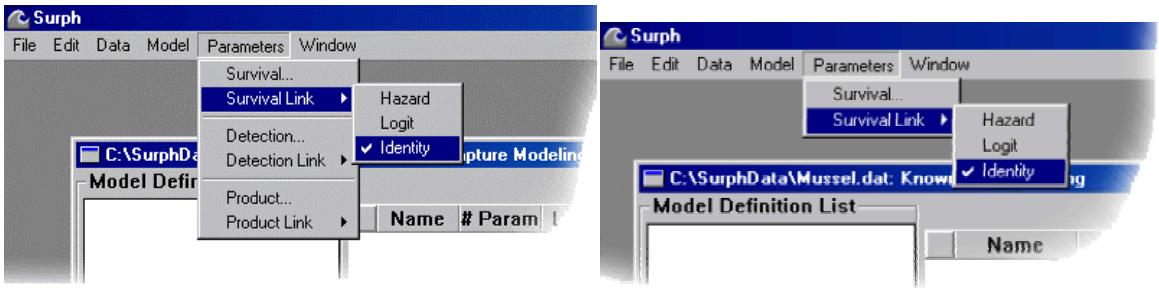
1. Set the link function and the parameter definitions for all SURPH probabilities, and encapsulate these settings by creating a model definition. This can be repeated to create multiple model definitions. Once created, the model definitions will appear in the Model Definition List of the SURPH Dialog Window.
2. Press the "Create Model(s)" button on the SURPH Dialog Window to estimate the parameters for all of the model definitions. The Model Definitions will disappear from the Model Definition List, and the corresponding Models will now appear in the Master Model List of the SURPH Dialog Window.
    - Once a model is created, the Calculate Variance option can be selected.

After the desired model(s) have been created, the user can:

- Display the output for a selected model.
- Perform hypothesis testing.
- Examine the model diagnostics.

4.1 Link Functions

Release-Recapture **Known-Fate**



The functional relationship of the model parameters to the SURPH survival probabilities is defined by the link function. SURPH 2.1 defines three link functions.

- The **identity link** assumes an additive effect on survival probability. The identity link is the default link for all SURPH probabilities.
- The **hazard link** assumes a multiplicative effect on hazard rate, an exponential effect on survival probability.
- The **logit link** assumes an additive effect on log odds ratio.

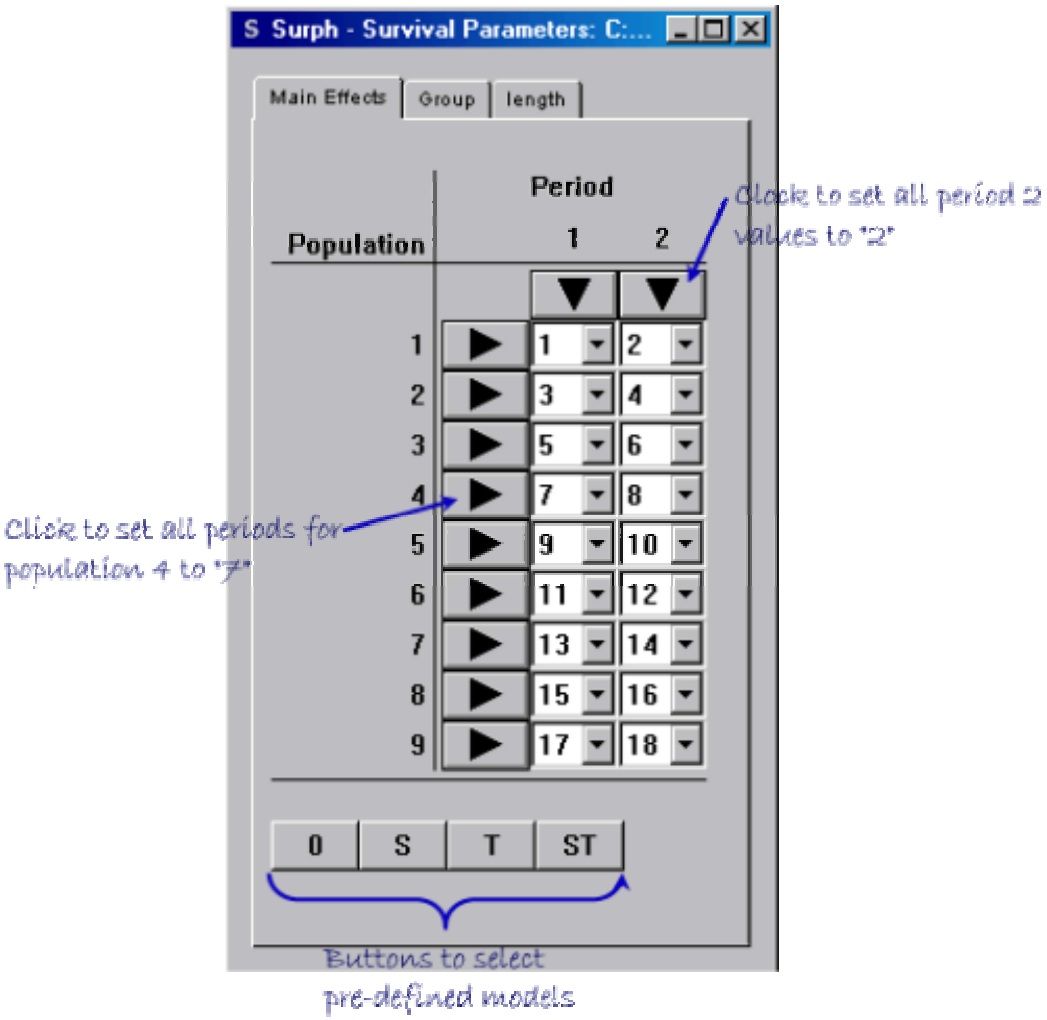
Both the hazard link and the logit link constrict the probabilities to be between 0 and 1; the identity link allows a probability to be greater than 1.

The same link function options are also available for modeling the detection probabilities and the joint product term in the last period. Link functions are specified separately for the survival, detection, and joint product terms.

4.2 Parameter Definitions

4.2.1 Main effects parameter definitions

The SURPH parameters for a given SURPH probability are set via the parameter dialog window. The parameter dialog shown below is the main effects parameter dialog window for a study with 9 populations and 2 periods (see Release-recapture notes).



For each period in each population, the user can select a number between 1 and N, where N is the number of populations times the number of periods. Each number represents the parameter for that population and period. If two selections are set to the same value, the corresponding populations and periods will share the same parameter.

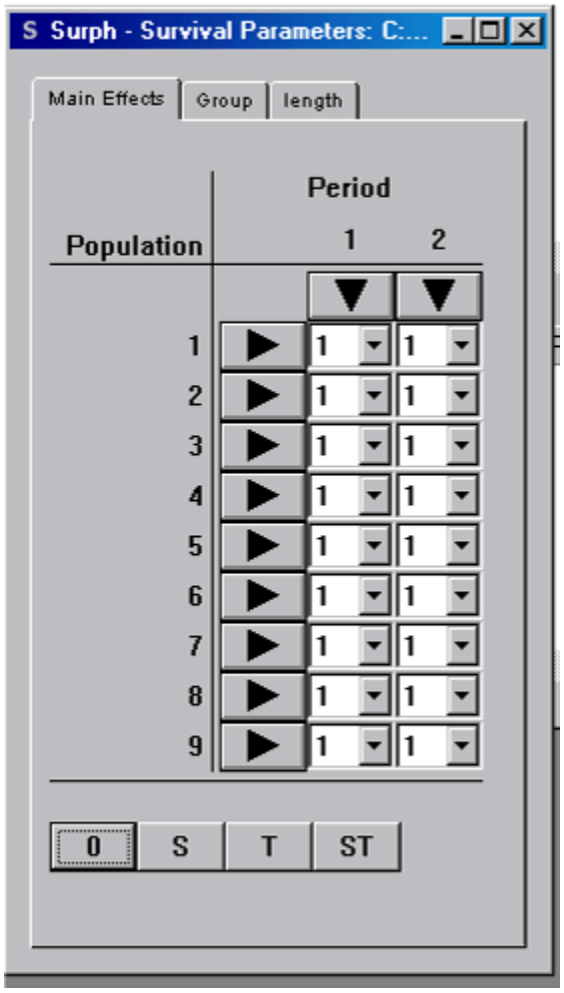
There are four pre-defined settings for the main effects dialog window:

Button	Model	Description
0	Null Model	There is one parameter across all populations and periods.
S	"Space" Model	There is a unique parameter for each population, but all periods share the same parameter.
T	"Time" Model	There is a unique parameter for each period, but all populations share the same parameters.
ST	"Space/Time" Model	There is a unique parameter for each population and each period. This is the default setting shown above.

Parameter sets for these four different model specifications are illustrated below along with an example of a user-specified model.

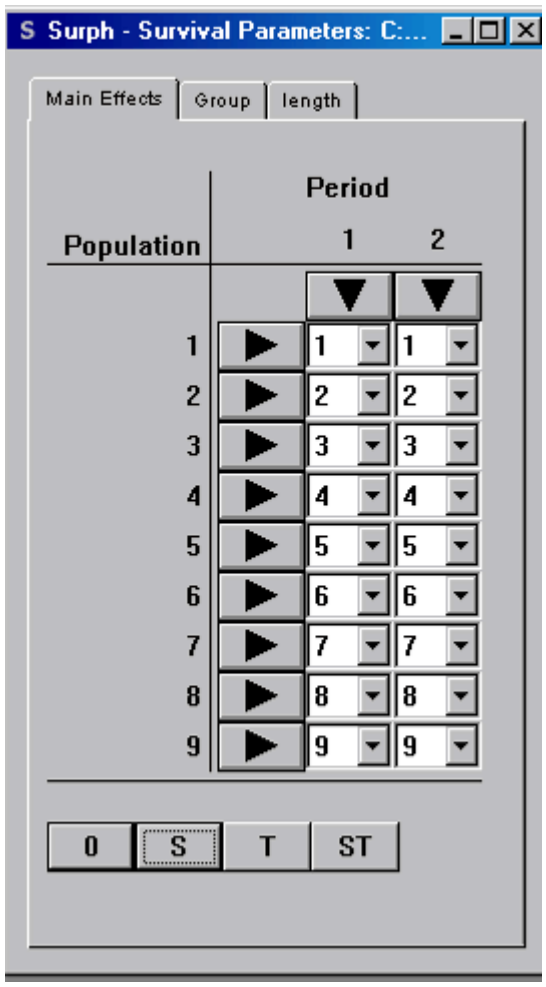
Null Model

The same parameter for all populations and periods



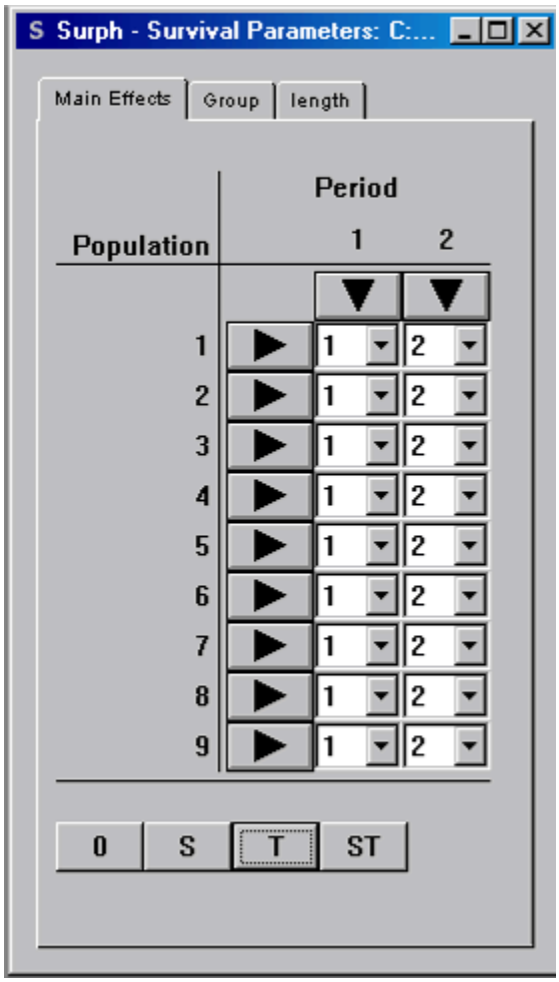
"Space" Model

Unique parameters for each population, shared parameters across periods



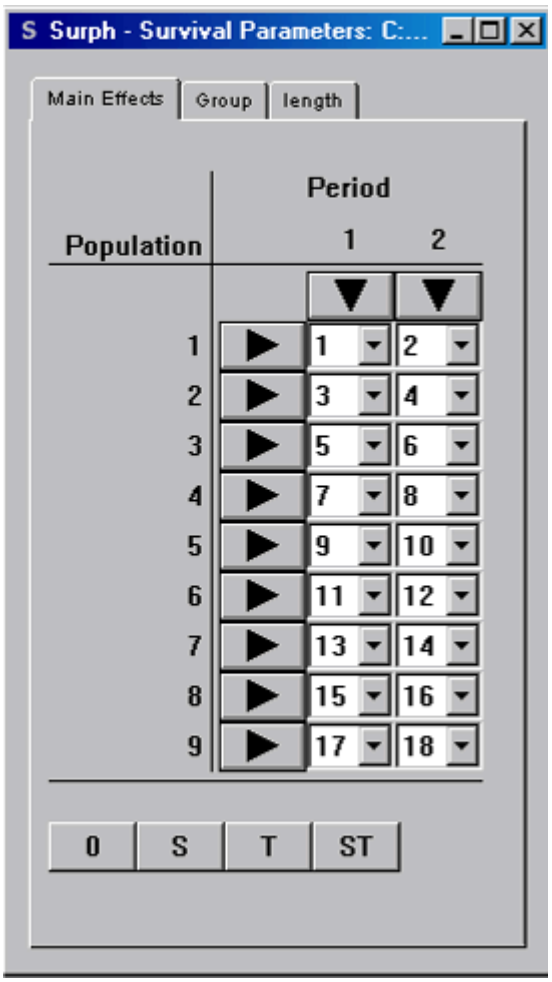
"Time" Model

Unique parameters for each period, shared parameters across all populations



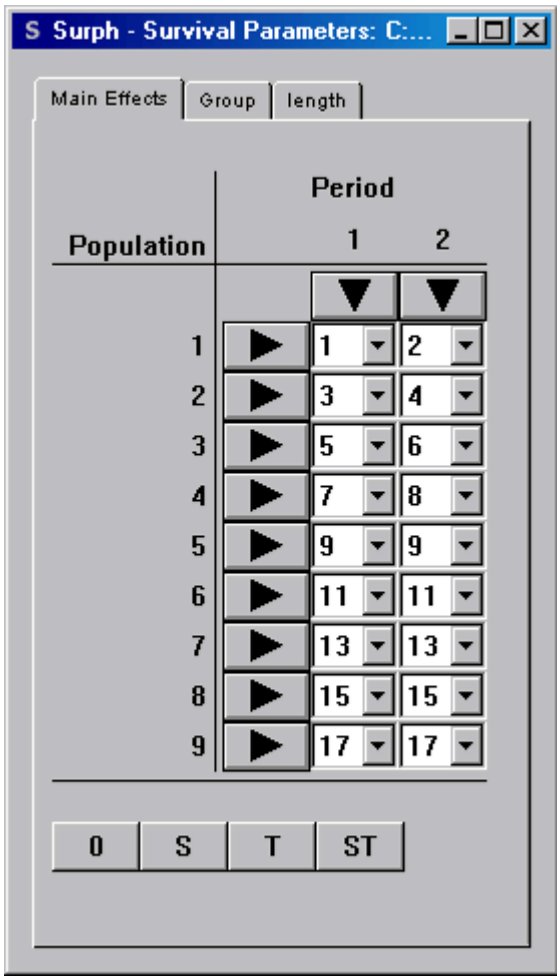
"Space/Time" Model

Unique parameters for each population and period



“Special Case” Model

First 4 populations: Unique parameter for each period and population;
Last 5 populations: Unique parameters across populations, shared parameters across periods.

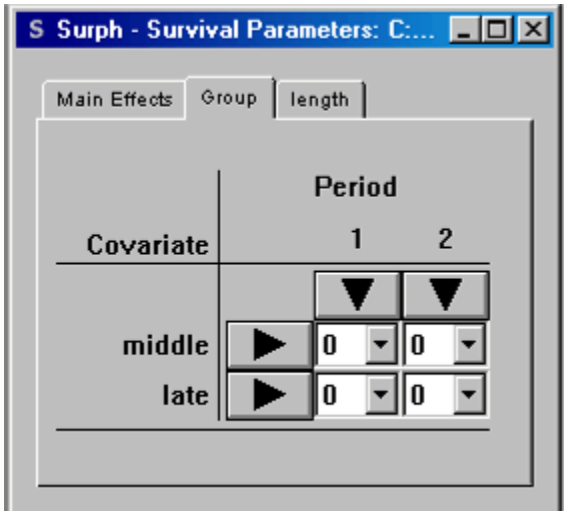


4.2.2 Group covariate parameter definitions

The group covariates for a SURPH probability are defined in much the same way as the main effect parameters, except that:

- There are no "predefined" model buttons;
- A value of "0" (zero) is allowed, indicating that the given group covariate is not modeled for the given population and period;
- Group covariates cannot be unique across populations – only across periods.

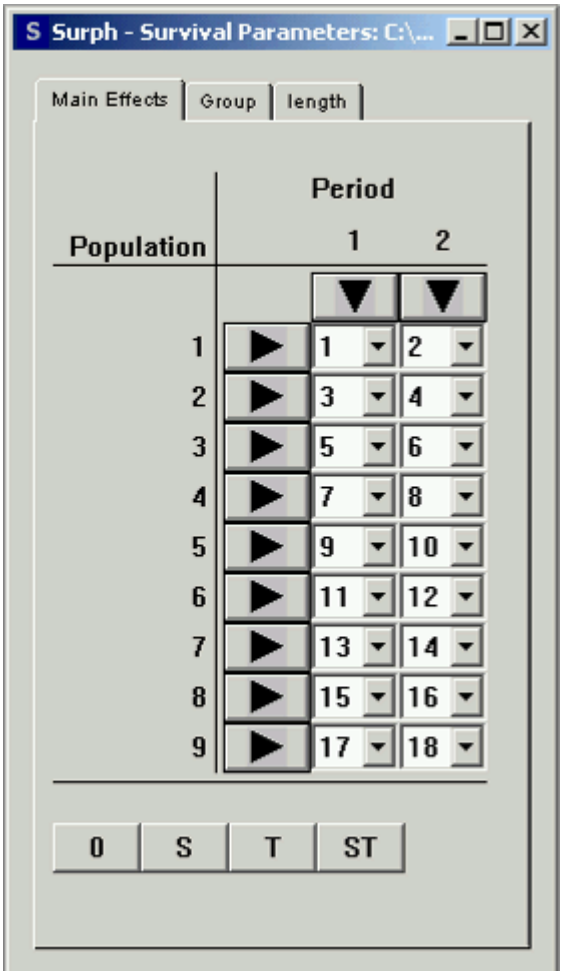
Below is an example of the group covariate dialog window for a study with 2 group covariates ("middle" and "late"), and 2 periods. It shows the default setting of all zeroes, indicating that no group covariates are included in the model (see Release-recapture notes).



As with the main effects dialog window , periods with the same number share a common parameter for a given covariate. Note, however, that parameters are not shared across covariates; each covariate has its own set of parameters.

4.2.3 Individual covariate parameter definitions

The individual covariate parameters are defined in a similar way as group covariate parameters. The parameter dialog window has a tab for each individual covariate. In the example below, there is one covariate "length" (see Release-recapture notes).



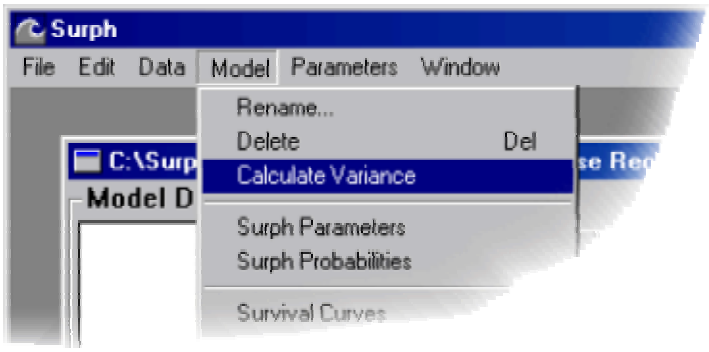
For a given individual covariate, a unique parameter can be specified for each population and period. The above example illustrates the default of all "0"s, specifying that the given individual covariate is not included in the model.

Again, the same number in multiple positions indicates a shared parameter.

4.2.4 Define model - Release-recapture notes

The model parameter dialog windows for the **product** SURPH probabilities (the product of the final detection and survival probability for each population) is different from the other SURPH probabilities in that there is no period effect for any of the parameters. With that one exception, the product probabilities may be modeled in the same way as the other SURPH probabilities.

4.3 Calculate Model Variance



When a model is created and the parameter estimates calculated, the variances of the parameter estimates are not automatically calculated, and the standard errors on the model output will be shown as a series of dashes ("-"). To obtain the variance estimates (and hence the standard errors), the user must explicitly calculate the parameter variances as shown above. If a model output report is open, it must be closed and reopened in order to show the updated standard errors.

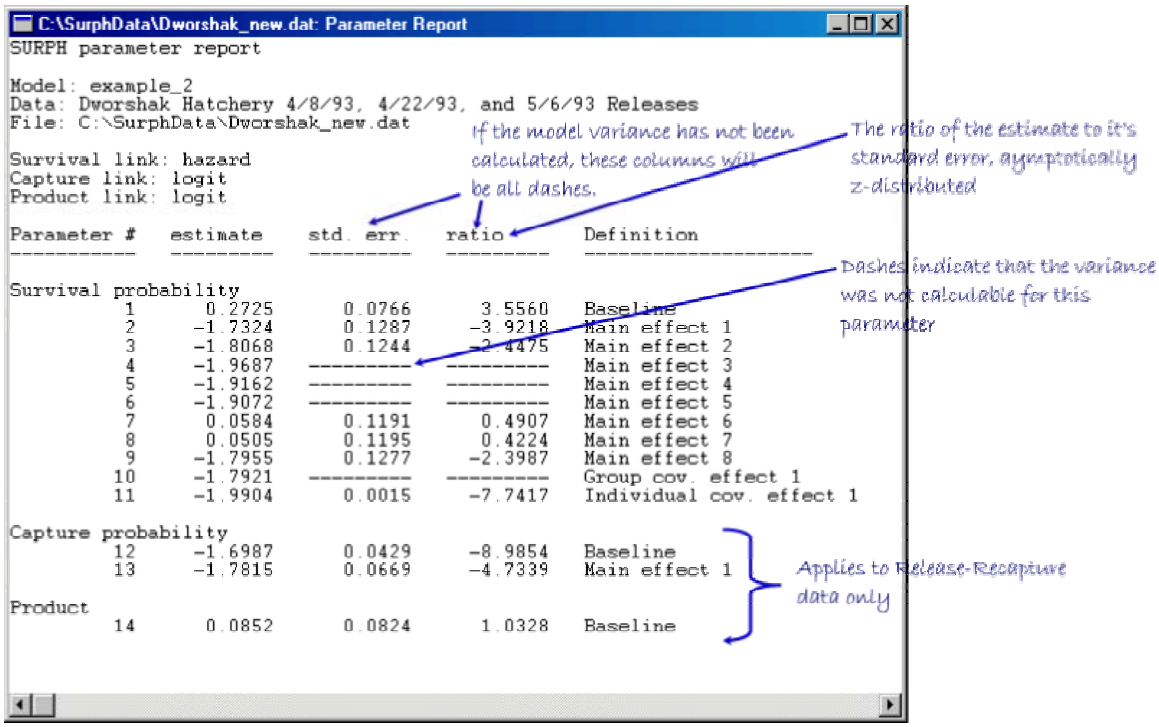
4.4 Model Output

There are two text reports for displaying the output of a selected model:

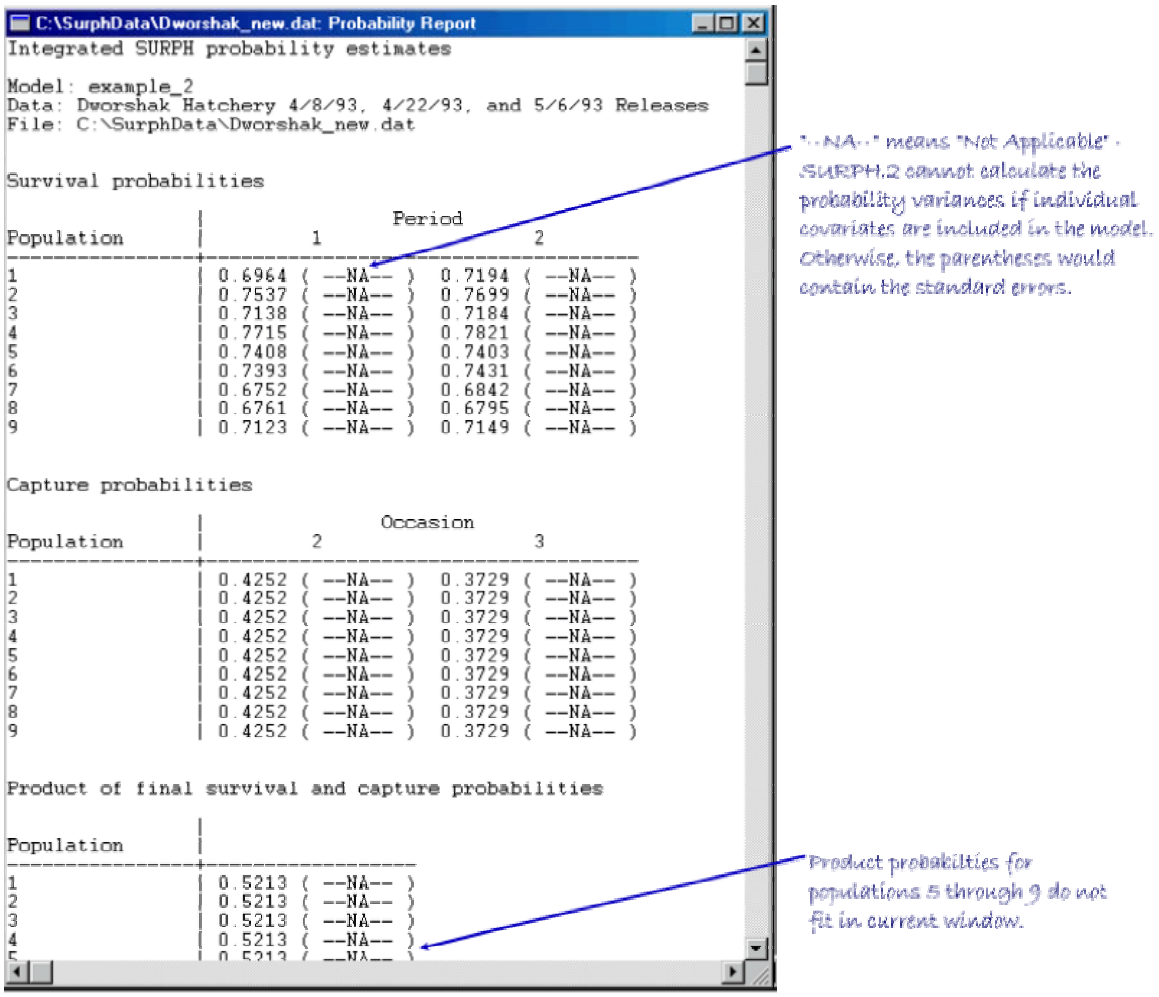
1. The SURPH Parameters Report shows the estimates of the model parameters with the standard error (if the model variance has been calculated).
2. The SURPH Probabilities Report shows the estimates of the SURPH survival and detection probabilities, calculated from the SURPH parameters with the selected link functions.

Examples of the two reports are show below.

4.4.1 SURPH parameters report



4.4.2 SURPH probabilities report



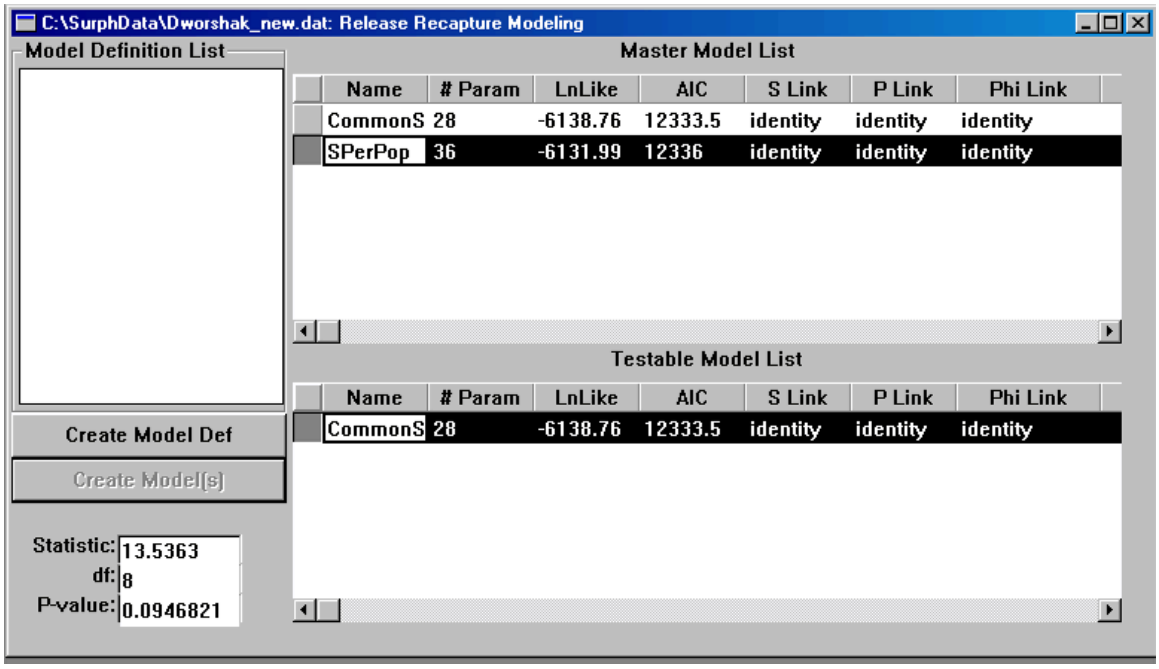
4.5 Hypothesis Testing

There are three ways of performing model selection and hypothesis testing within SURPH:

1. Likelihood Ratio Test for testing nested or hierarchical models.
2. Akaike Information Criteria for non-nested models.
3. Analysis of Deviance for testing group covariate effects.

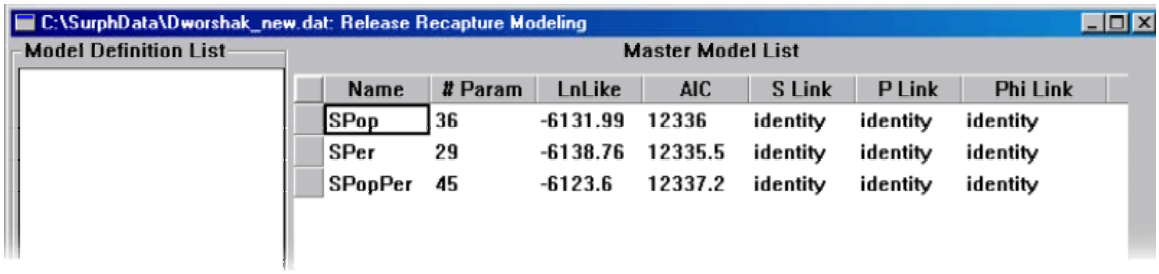
4.5.1 Hypothesis testing - Likelihood ratio test

The Likelihood Ratio Test (LRT) is a general, powerful method of testing two models when one model is a nested or special case of the other. In the illustration below, "CommonS" is a special case of "SPerPop" and therefore is displayed in the Testable Model List when "SPerPop" is selected in the Master Model List. If "CommonS" is then selected in the "Testable Model List", a Likelihood Ratio Test is performed, with the test results appearing in the lower left of the SURPH Dialog Window.



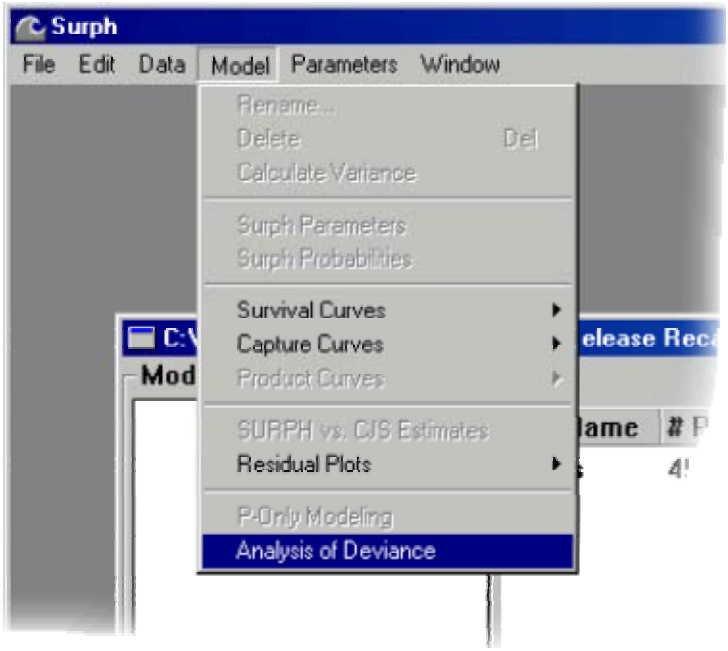
4.5.2 Hypothesis testing - Akaike information criterion

The Akaike Information Criterion, or AIC (Akaike 1973, Sakamoto et al. 1986), has received much recent attention in release-recapture literature (e.g., Burnham and Anderson 1998). The details of the AIC are provided in the SURPH.1 manual. The smaller the value of the AIC, the better the fit of the model to the observed data. The AIC, unlike the Likelihood Ratio Test, has been used to select among non-nested models. In the example below, the model "SPer" has the smallest AIC (12335.5) and would be selected as the better fit based on the AIC.



Name	# Param	LnLike	AIC	S Link	P Link	Phi Link
SPop	36	-6131.99	12336	identity	identity	identity
SPer	29	-6138.76	12335.5	identity	identity	identity
SPopPer	45	-6123.6	12337.2	identity	identity	identity

4.5.3 Hypothesis testing - Analysis of deviance



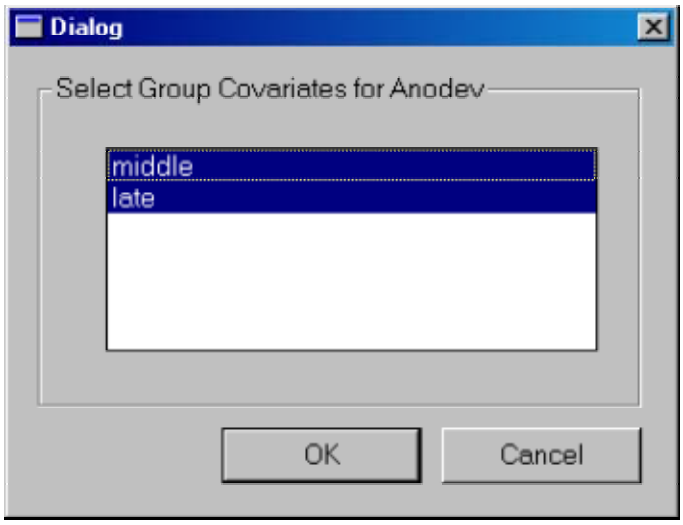
Simulation studies have shown that when data were generated with no group covariate effects between populations (the null hypothesis), the Likelihood Ratio Test

consistently rejected the null hypothesis far too often (Smith 1991). The Analysis of Deviance (ANODEV) approach, however, performed according to its expected nominal distribution (Smith 1991). For this reason, SURPH 2.1 provides an Analysis of Deviance procedure for testing models with group covariate effects.

The theory behind Analysis of Deviance and a complete description of the models can be found in the SURPH.1 Manual; the following is simply a description of the procedure for using the Analysis of Deviance with SURPH 2.1.

4.5.3.1 Analysis of deviance - Initialization

Select the Covariates



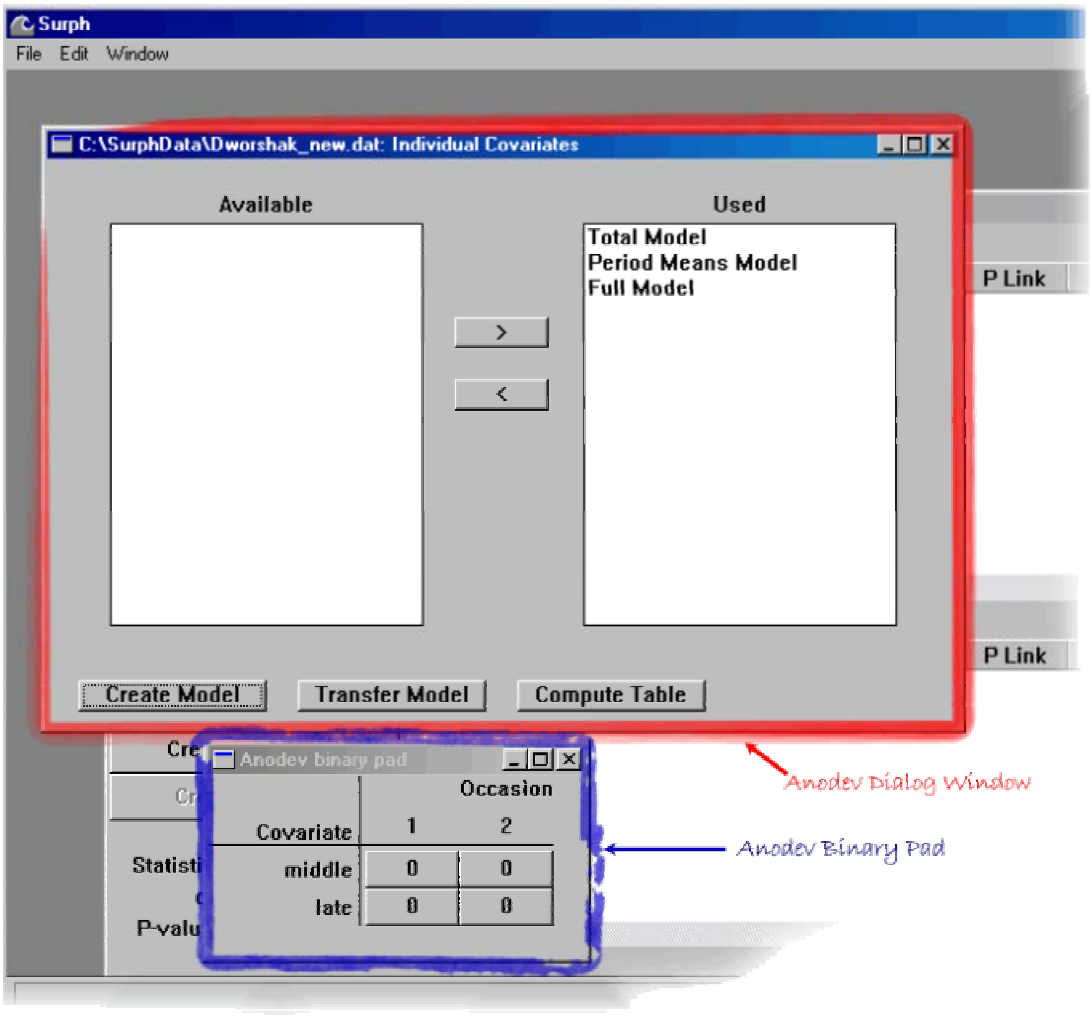
After the Analysis of Deviance option is selected, the above dialog window appears to allow the user to select which group covariates will be used in the Analysis of Deviance modeling. Multiple covariates may be selected by holding the "cntrl" key or the "shift" key when selecting with the mouse.

Anodev Dialog Window and Binary Bad

Once the user has selected the covariates to be used, SURPH 2.1 creates the three default Anodev models:

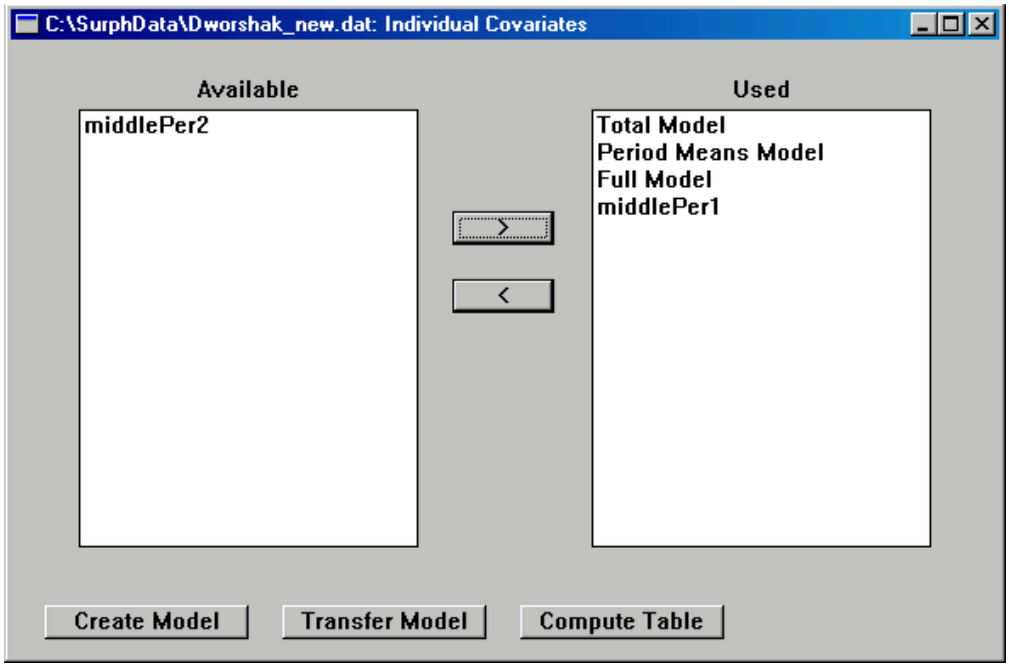
1. Total Means Model.
2. Period Means Model.
3. Full Model.

The Anodev Dialog Window and the Anodev Binary Pad then appear, as show below, with the three default models in the Used list.



4.5.3.2 Analysis of deviance - Components

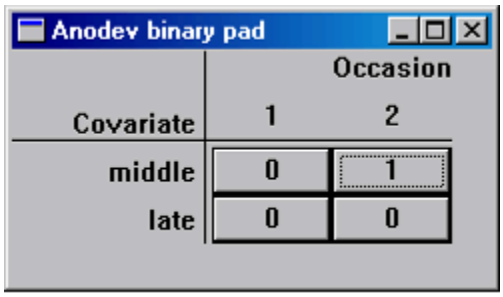
Anodev Dialog Window



The Anodev Dialog Window is the window that controls the Analysis of Deviance procedure. Its components are:

Available	The list of all models that have been created by the user during the current Analysis of Deviance session. It does not include the three default models created during initialization.
Used	The list of all models used in computing the Anodev Table, created with the Compute Table button.
Create Model	Creates a new model based on the current setting of the Binary Pad. The new model appears in the **Available** list.
Compute Table	Recalculates and displays the Anodev Table based on the models currently in the **Used** list.
>	Adds the model selected in the **Available** list to the **Used** list. An error message will be displayed if the model cannot be added because it is not nested in relation to the other models in the **Used** list.
<	Removed the selected model from the **Used** list. The model is placed back into the **Available** list.

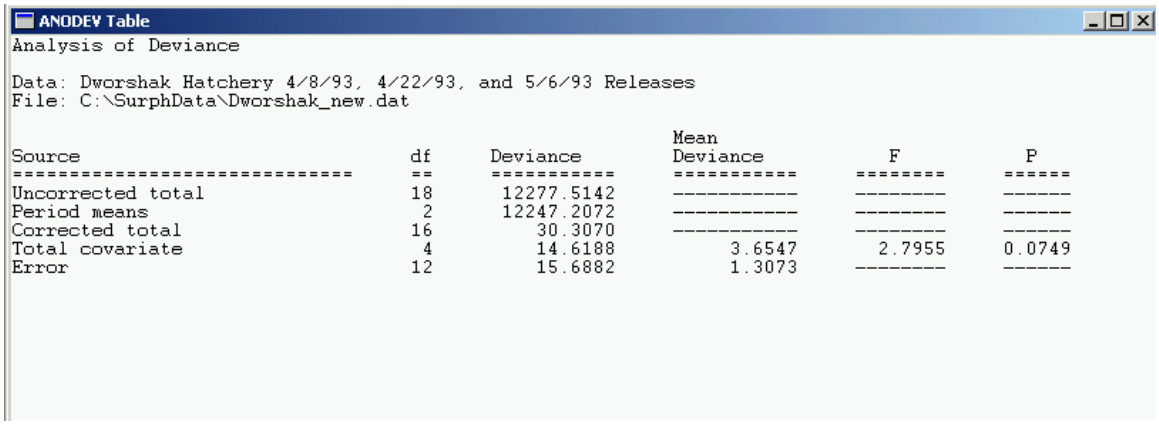
Anodev Binary Pad



The Anodev Binary Pad functions in much the same way as a SURPH Parameter Dialog Window, except that the options for each covariate and period are limited to "0" to disable the parameter for the given covariate and period, or "1" to enable it. The value of a button is toggled by clicking on it. In the example shown above, pressing the "Create Model" button will create a model that includes the covariate "middle" in the 2nd period only, and does not include the covariate "late" at all.

Anodev Table Report

When the "Compute Table" button is pressed on the Anodev Dialog Window, the Anodev Table Report appears (or is updated if already present), showing the Anodev Table for the models in the **Used** list of the Anodev Dialog Window.



Source	df	Deviance	Mean Deviance	F	P
Uncorrected total	18	12277.5142			
Period means	2	12247.2072			
Corrected total	16	30.3070			
Total covariate	4	14.6188	3.6547	2.7955	0.0749
Error	12	15.6882	1.3073		

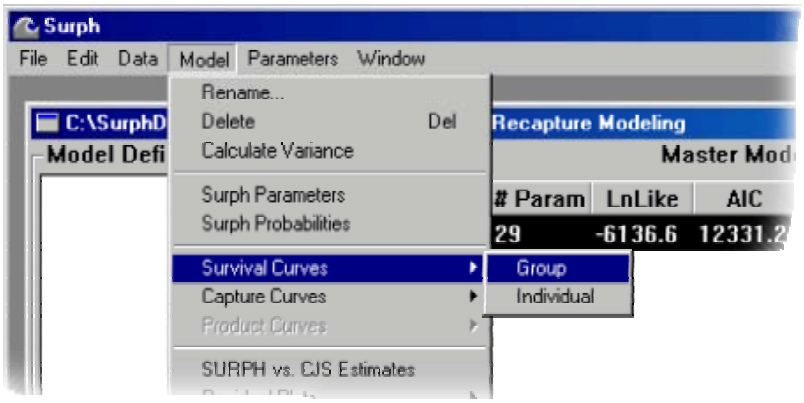
4.6 Model Diagnostics

SURPH 2.1 provides tools for examining how well a given model fits the data. They are:

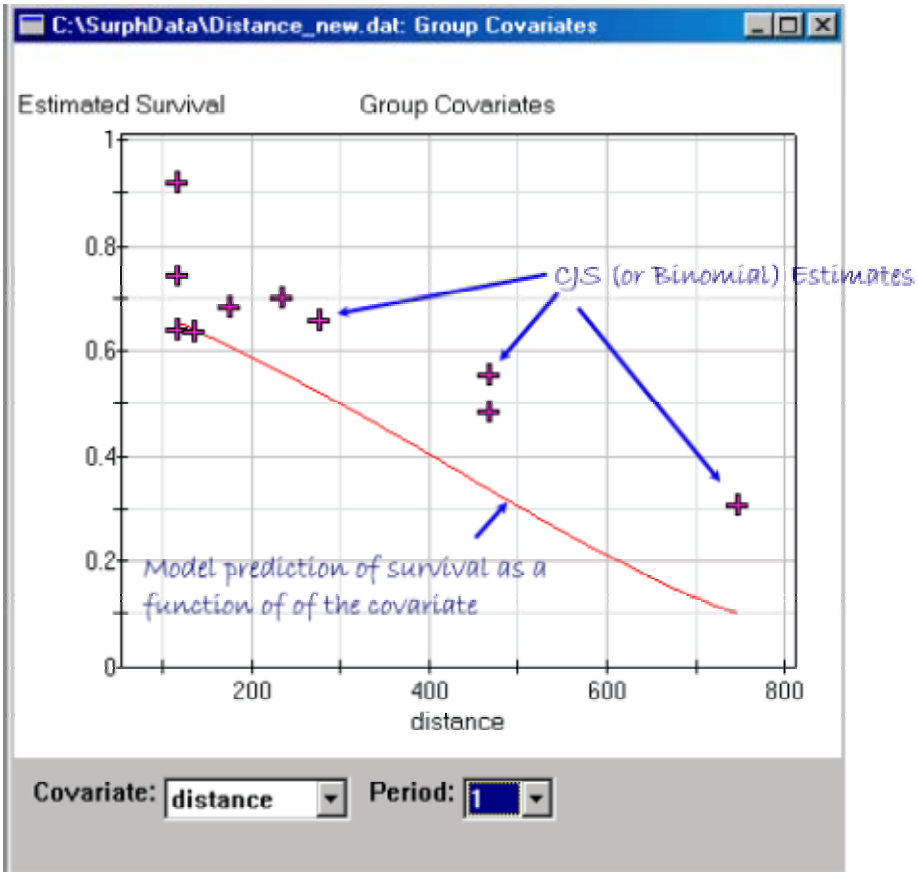
- Group covariate survival curves and individual covariate survival curves that graphically illustrate how the model maps a given SURPH probability to the given covariate (for release-recapture data, note that all of the same survival curve functionality is available for detection probabilities as well).
- For release-recapture data, a graphical comparison of the CJS estimates to the model estimates of survival and detection probabilities.

SURPH 2.1 allows a user to display and evaluate the estimated survival as a function of a given group covariate or individual covariate. For release-recapture data, note that all of the same survival curve functionality is available for detection probabilities as well.

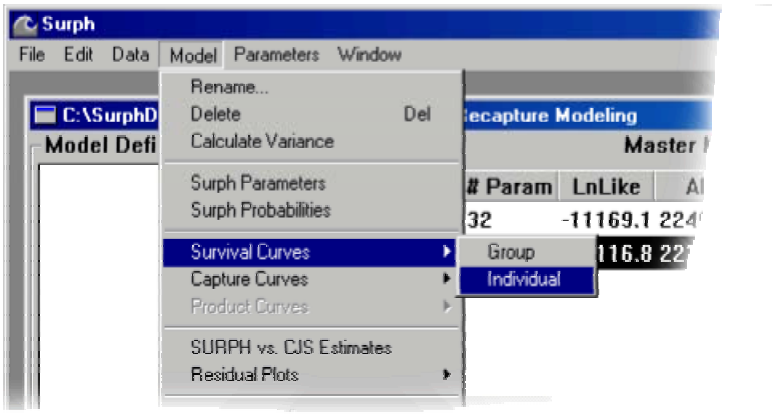
4.6.1 Model diagnostics - Group covariate survival curves



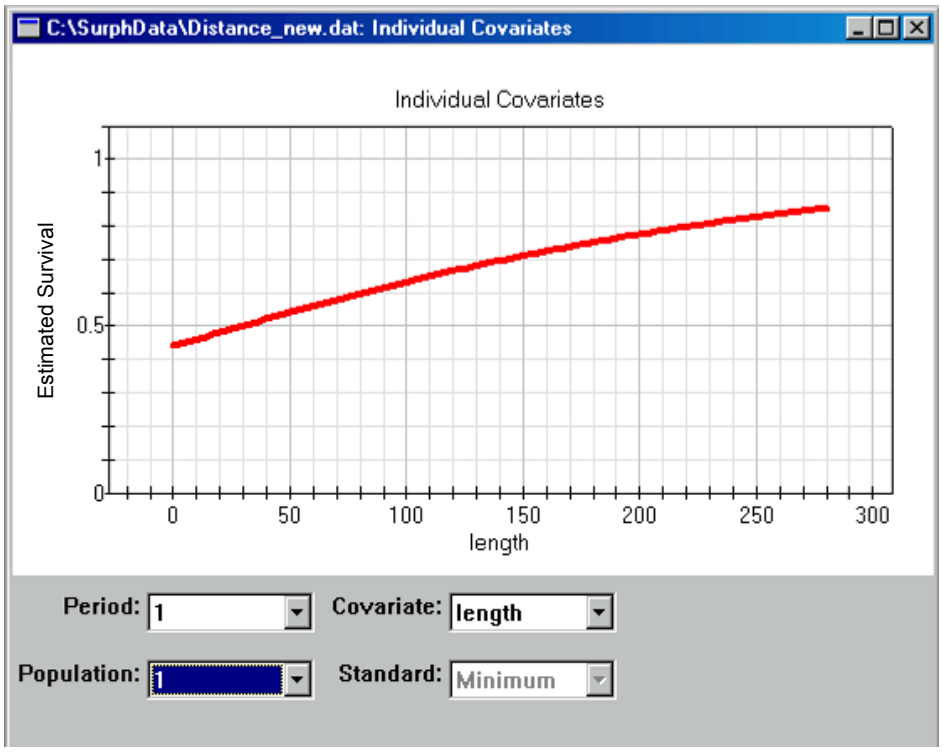
The group covariate survival curves allow the user to asses how well the given model's prediction of survival as a function of the covariate (the red line in the graph below) fits the actual data (the "+"s below).



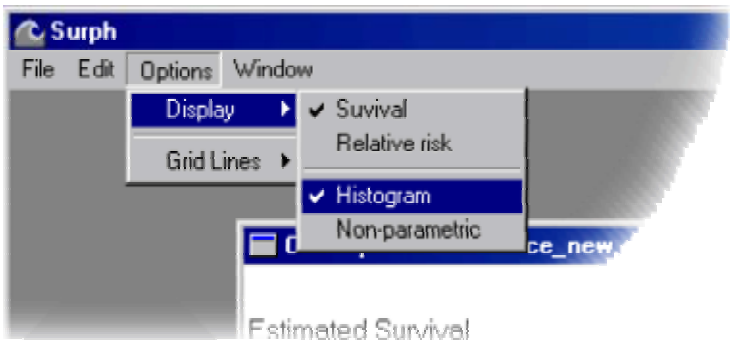
4.6.2 Model diagnostics - Individual covariate survival curves

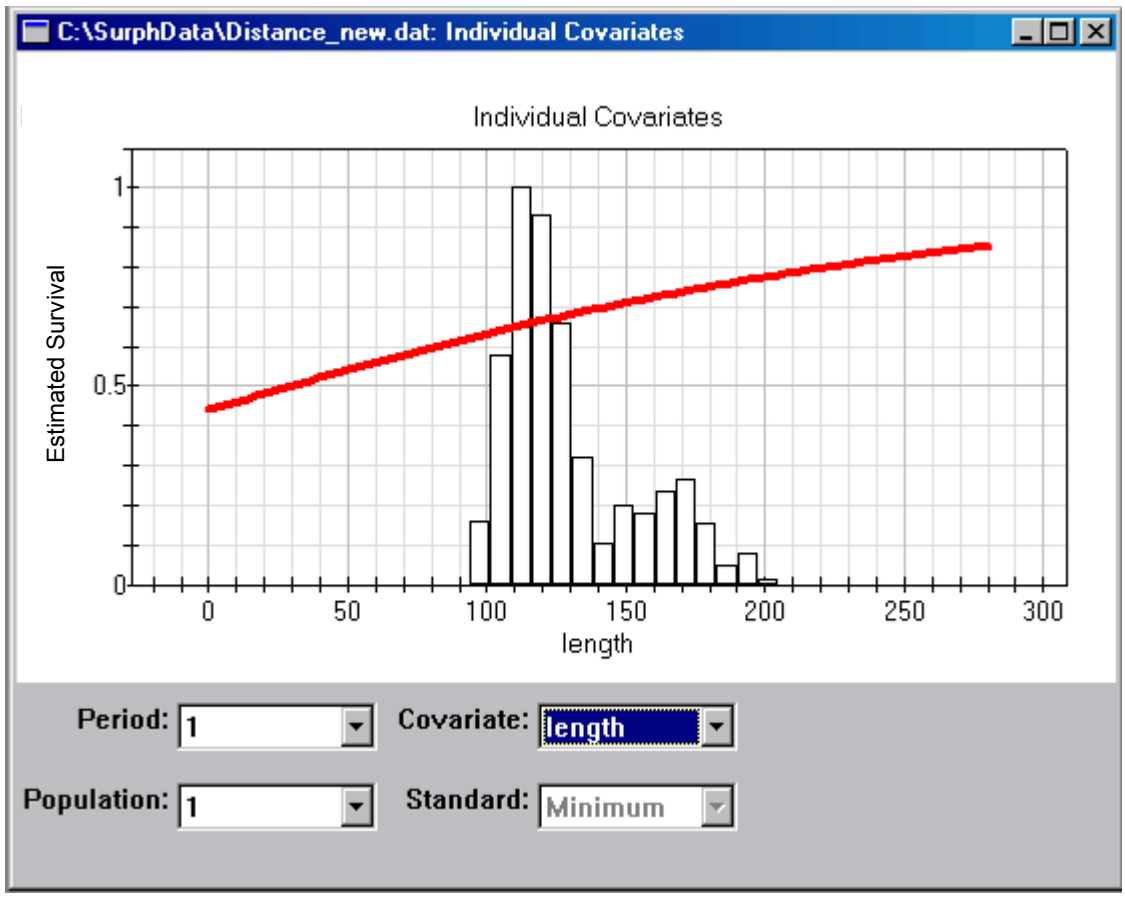


The individual covariate survival curves allow the user to assess how well a given model's prediction of survival as a function of the covariate (the red line in the graph below) fits the actual data. The user has the option of superimposing either a histogram of the data, or a nonparametric estimate of survival as a function of the covariate, over the model's prediction of survival as a function of the covariate.

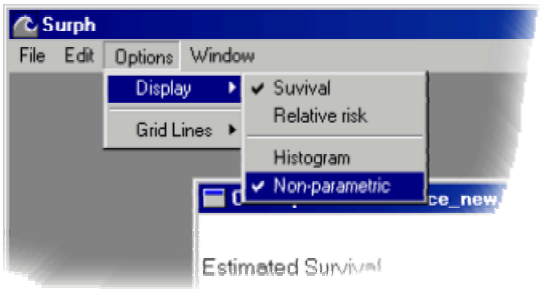


4.6.2.1 Histogram



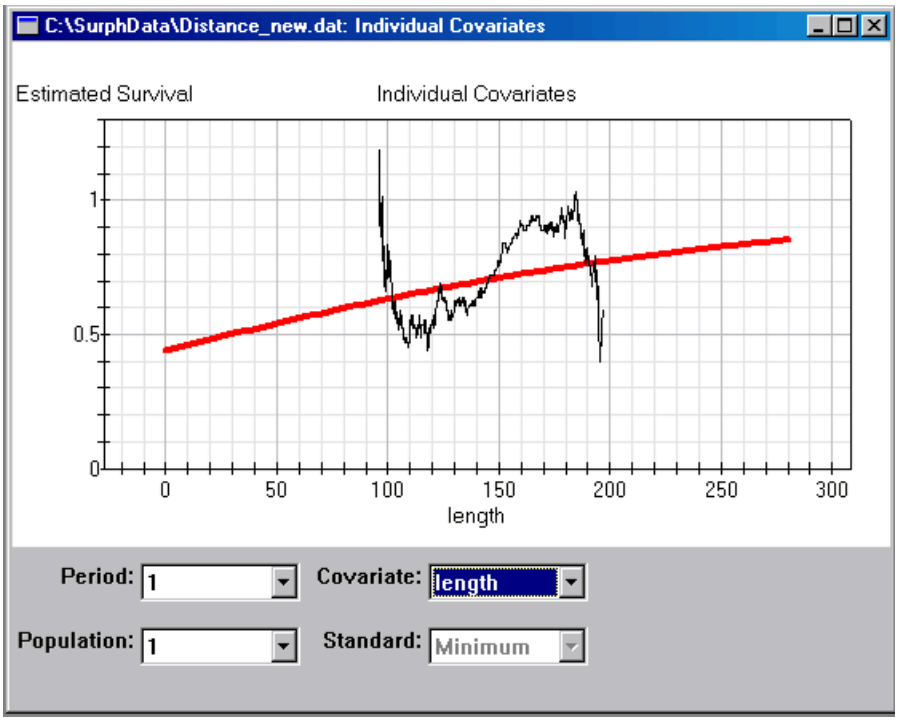


4.6.2.2 Nonparametric estimate

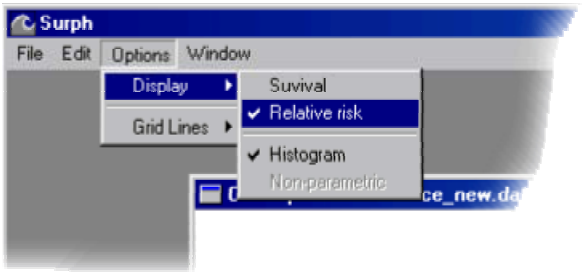


The nonparametric estimate of survival as a function of an individual covariate is superimposed on the model's prediction of the same function in the graph below. Each point on the nonparametric function is the CJS estimate (or binomial estimate for known-fate data), for all data with covariate values within a window around the given point. The window size varies from a minimum of eight data points up to one-fifth of the total number of data points.

Because the window of data used to create the separate CJS estimates is small, there is considerable variability about the individual estimates. Furthermore, the estimates from the moving window are correlated, resulting in localized trends that may not actually exist. Also, the window size is smaller at the lower and upper boundaries of the covariate.



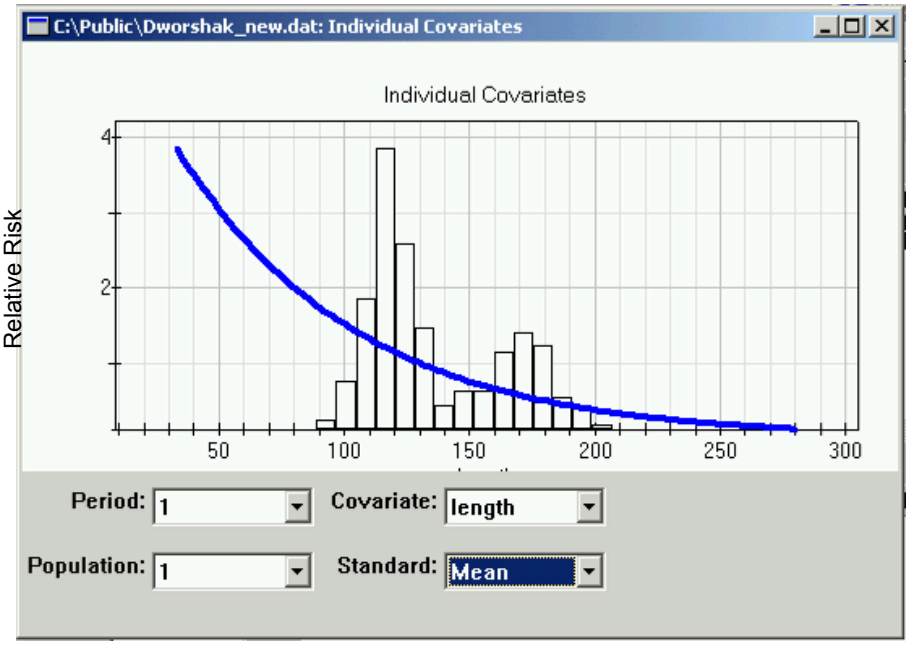
4.6.2.3 Relative Risk



If the Relative Risk option is selected from the graph options menu, the instantaneous relative risk (described in the SURPH.1 manual) as a function of the individual covariate is plotted instead of the survival function. As with the survival function graph, the user may superimpose a histogram of the covariate data on the graph.

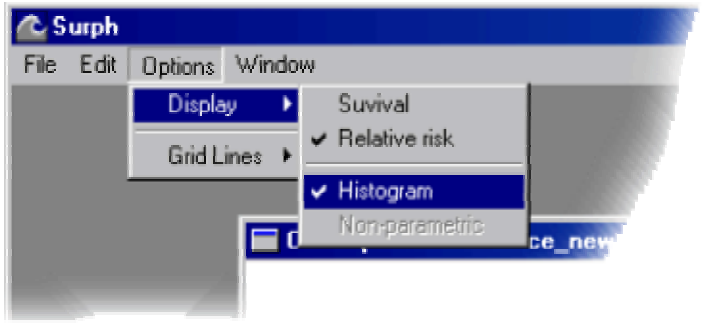
The user can select from one of three standards for the relative risk graph:

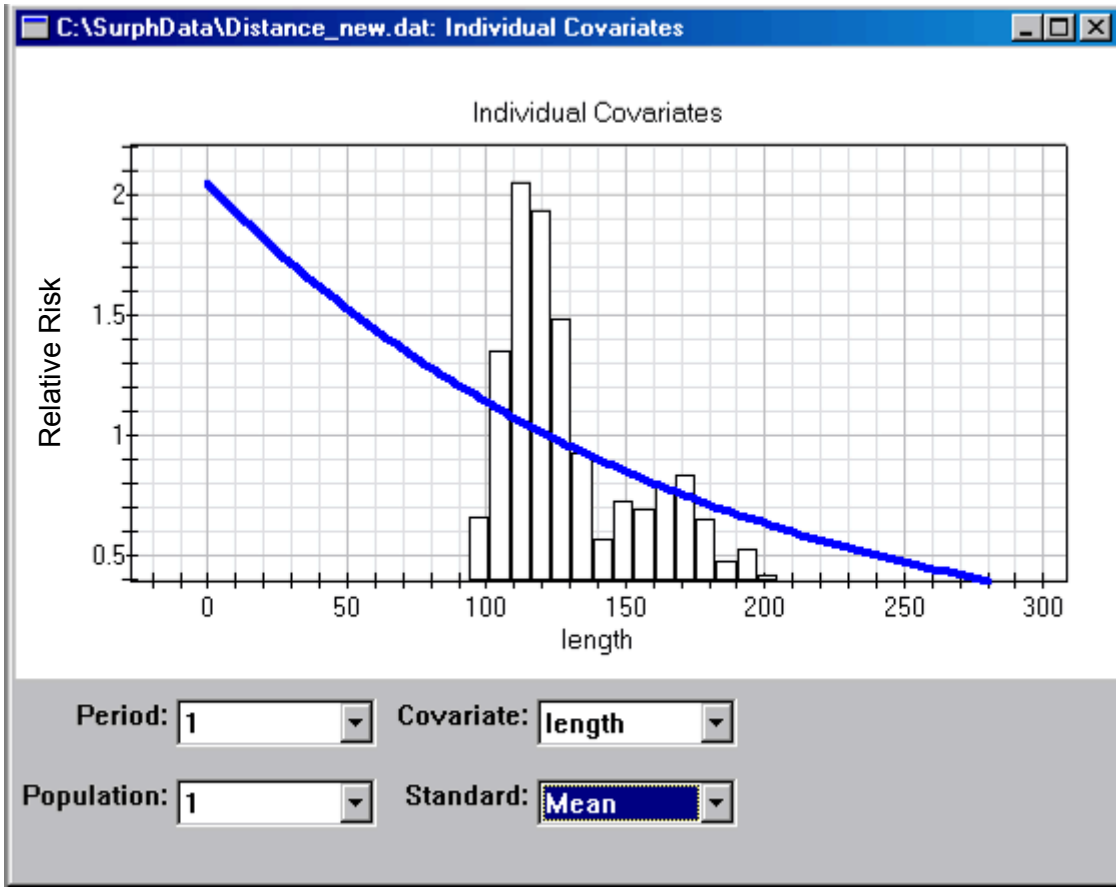
Mean	Relative risk of 1.0 corresponds with the mean covariate value (the default, as shown below).
Minimum	Relative risk of 1.0 corresponds with the minimum covariate value.
Maximum	Relative risk of 1.0 corresponds with the maximum covariate value.



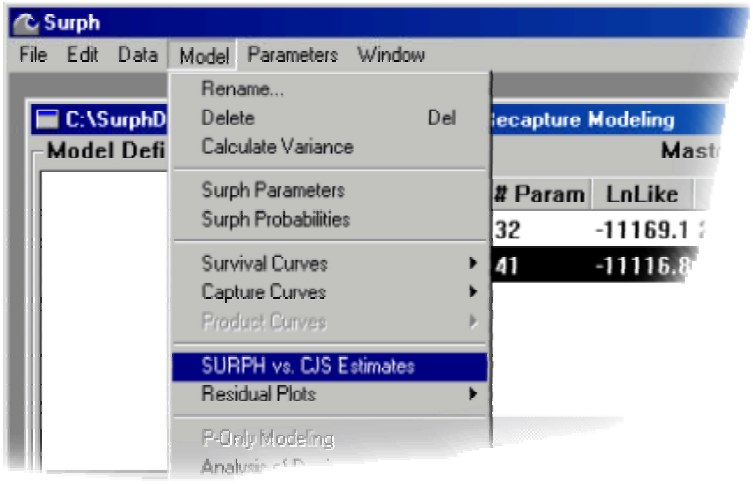
This relative risk curve is only applicable when the hazard function is used to model survival probabilities.

4.6.2.4 Relative Risk with Histogram

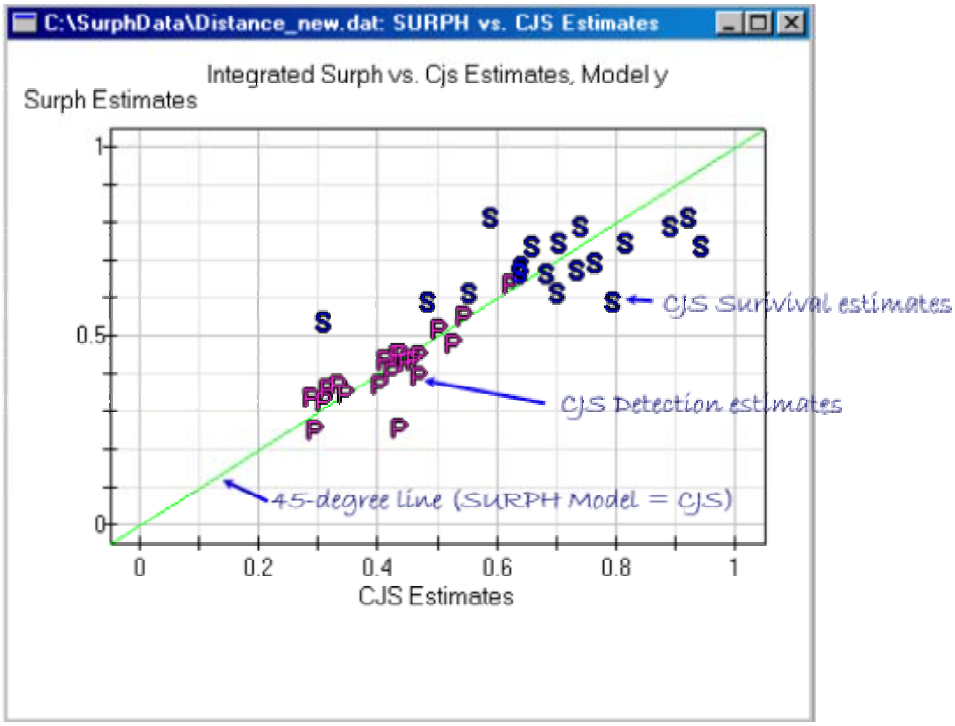




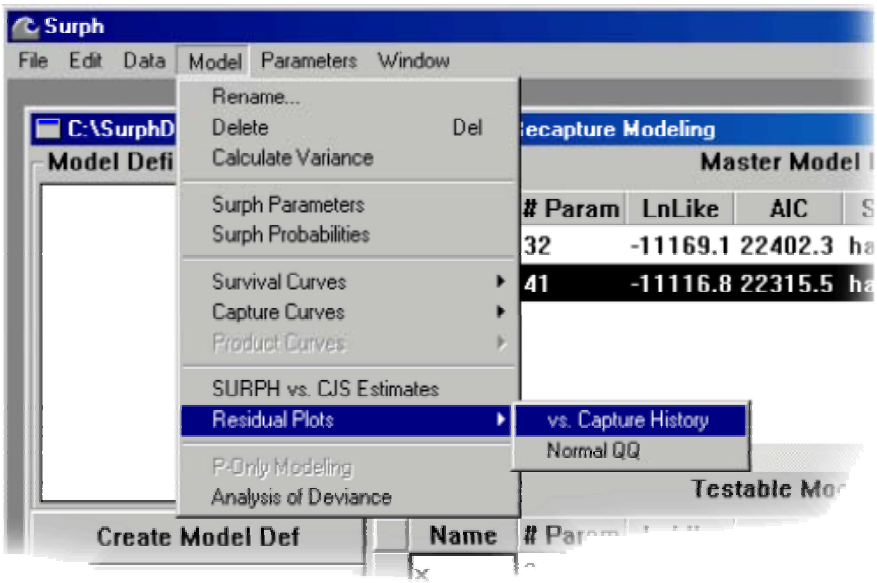
SURPH vs. CJS estimates



The Cormack-Jolly-Seber Estimates are unique for each population and period, and therefore fit the data optimally and provide a standard for evaluating the survival estimates from a selected model. The SURPH vs. CJS Estimates plot is illustrated below. The x-axis is for the CJS estimates, and the y-axis is for the SURPH estimates for the selected model. The green 45-degree line corresponds to the optimal condition of the two sets of estimates being equal; the distance from the green line corresponds to the discrepancy between the two estimates.

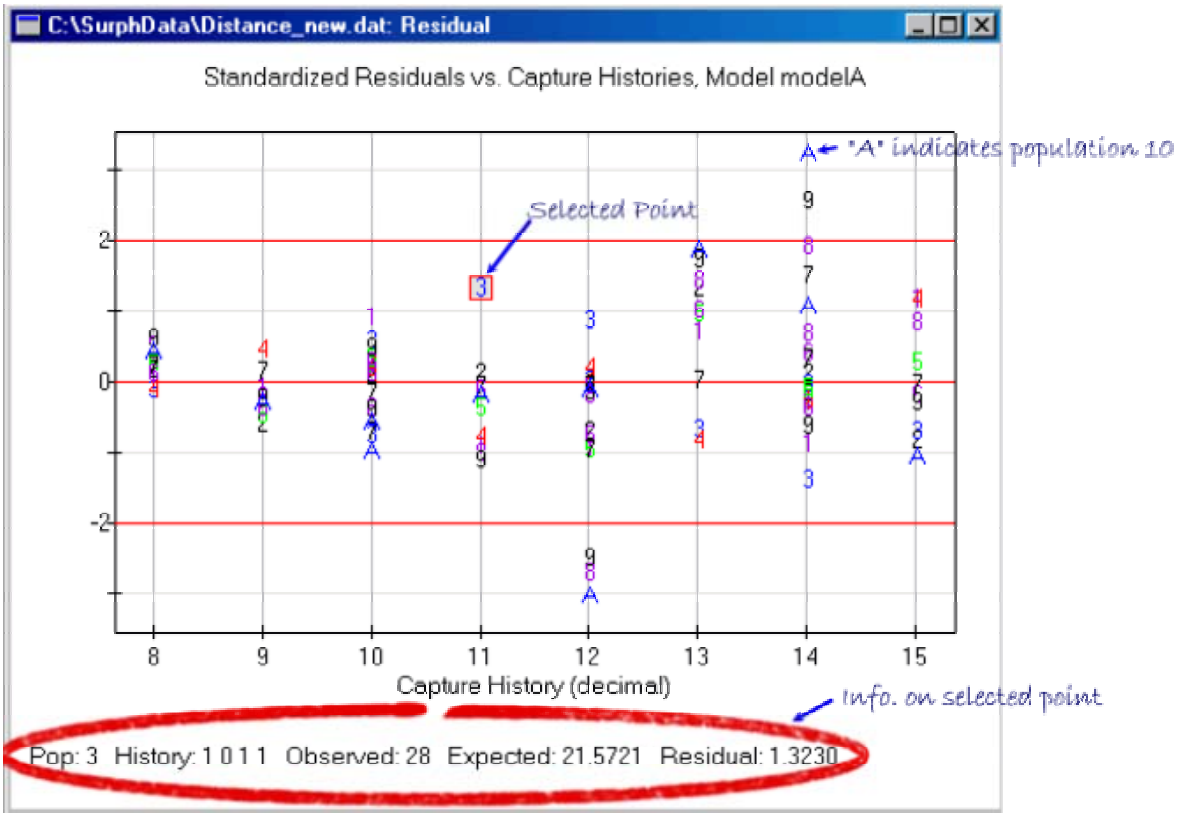


4.6.3 Residuals vs. capture histories

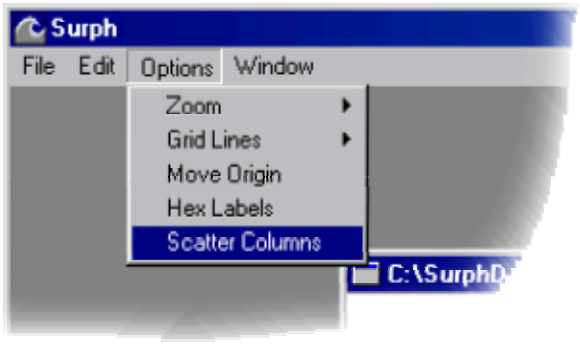


The Residuals vs. Capture History plot shows all of the Anscombe residuals for each possible capture history. The Anscombe residuals are the residuals transformed so as to follow a normal distribution (McCullagh and Nelder 1989). Each residual is indicated by a number indicating which population it is from, with "A" used for population 10, "B" for 11, etc. Horizontal red lines indicate values of -2.0 and +2.0. If the model adequately fits the data, the Anscombe residuals should follow a normal distribution, and approximately 95% of them should lie within the two horizontal red lines.

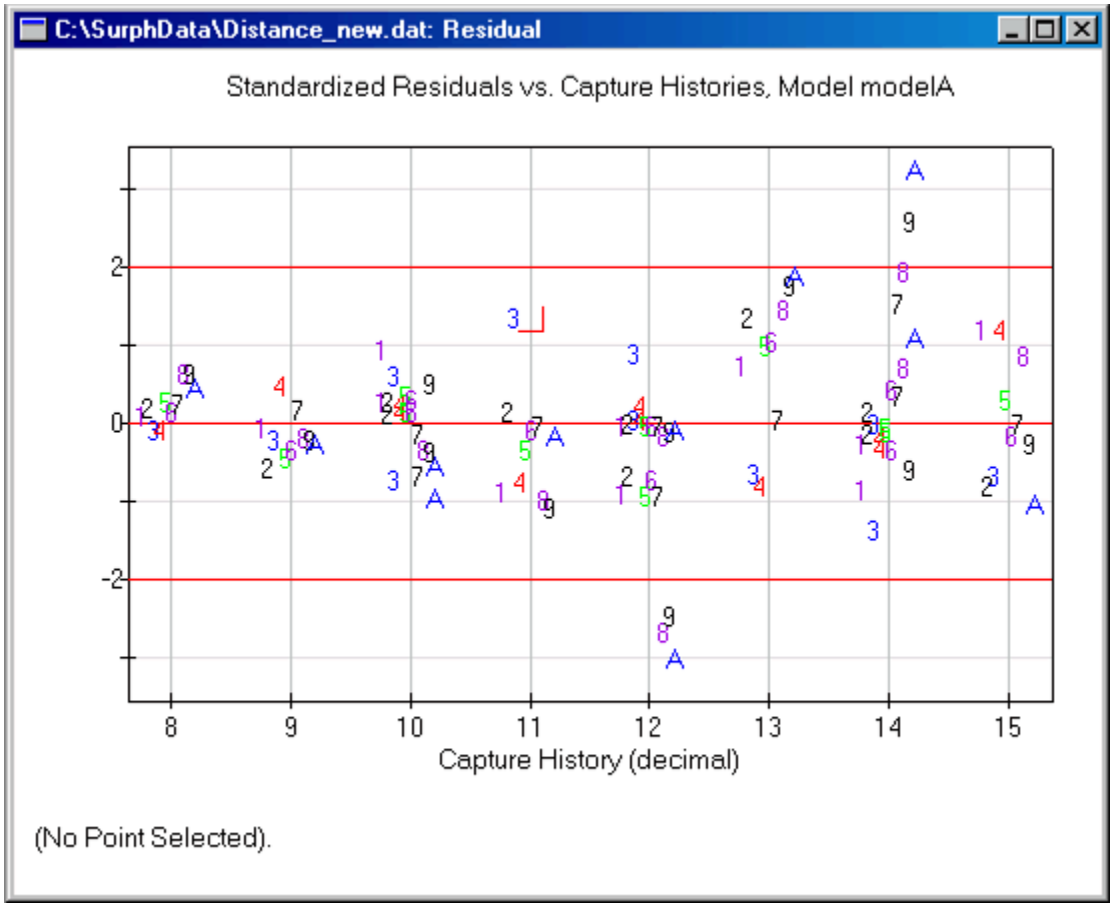
The user can select any data point on the plot, and the information for that point will be displayed along the bottom, as shown below. The user can also select the Scatter option to see the individual points more clearly.


C:\SurphData\Distance_new.dat: Residual
Standardized Residuals vs. Capture Histories, Model modelA
"A" indicates population 10
Selected Point
2
0
-2
8
9
10
11
12
13
14
15
Capture History (decimal)
Info. on selected point
Pop: 3 History: 1 0 1 1 Observed: 28 Expected: 21.5721 Residual: 1.3230

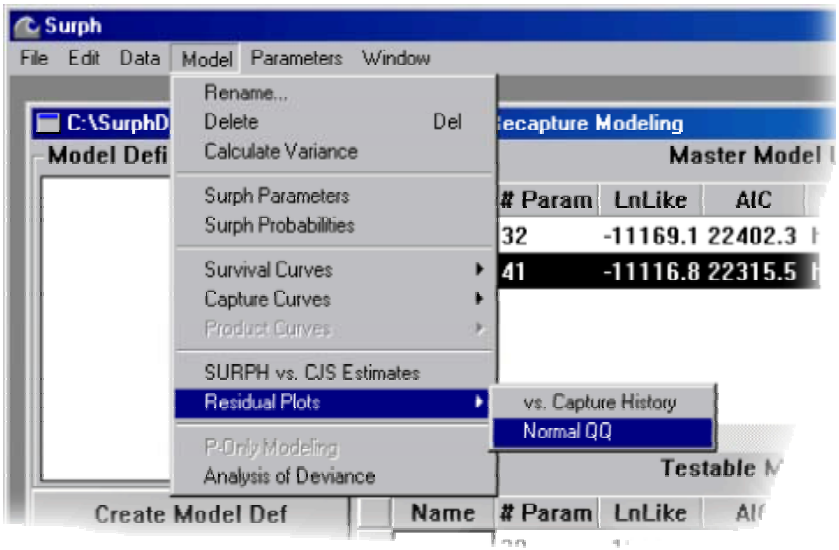
4.6.4 Scatter option



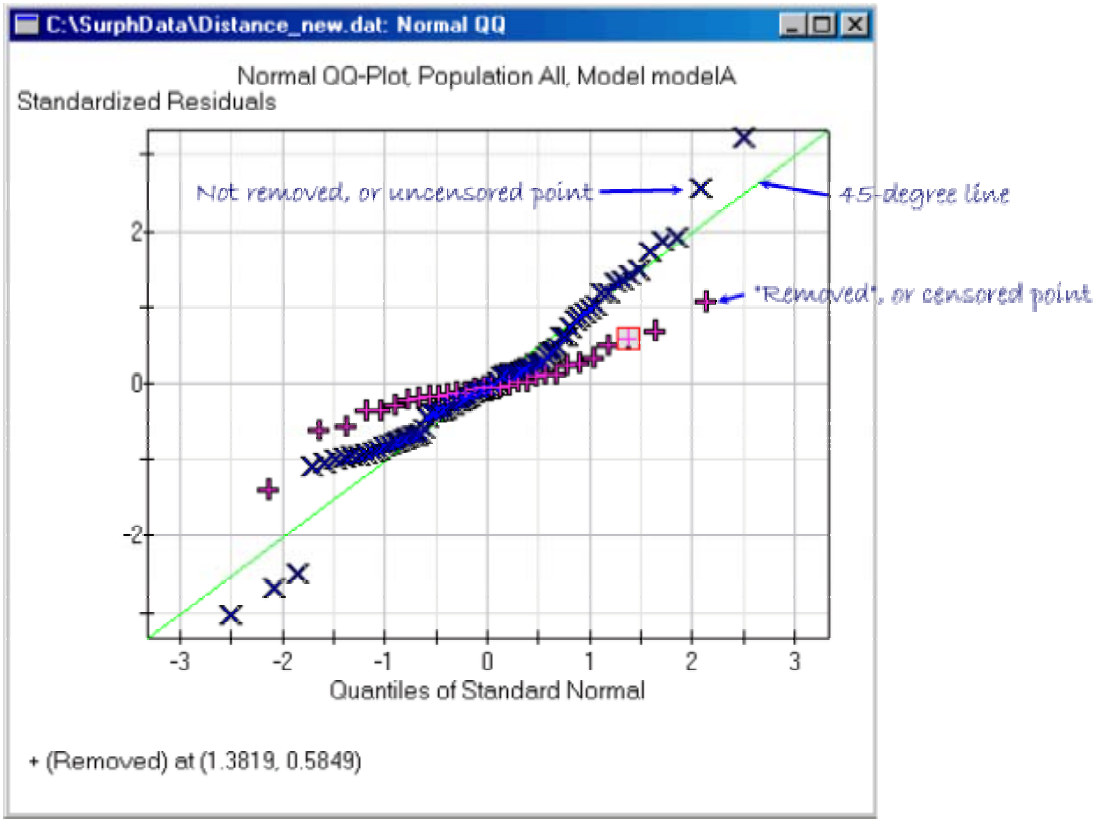
The scatter option offsets the points for the capture histories around the columns for the capture history so that the individual points may be more easily seen and selected, as shown below.



4.6.5 Normal QQ plot



The Normal-QQ plot is a plot of values of the ordered standardized Anscombe residuals. If the residuals follow a normal distribution, they should lie approximately on the green 45-degree line.



4.7 File Formats

The input data file is a text file that contains the following items in the given order.

1. The text "Surph2" on the first line
2. A one-line data description
3. The number of populations
4. The number of periods, or intervals
5. Group covariate definitions (if any)
6. Individual covariate definitions (if any)
7. Population names (optional)
8. The number tagged in each population
9. The Tag ID flag
10. A capture history line for each individual

SURPH.1 Users: Use the stand-alone data conversion routine to convert an input data file from SURPH.1 format to SURPH 2.1 format.

4.7.1 Input file

The first line of the input file must read “SURPH 2.”

4.7.2 Data description

The data description goes on the 2nd line (after the "Surph2" keyword). This description will appear in all reports based on the data.

4.7.3 Number of populations

The number of populations is indicated by the keyword **npop** or **numPopulations** followed by an integer for the number of populations.

4.7.4 Number of periods

The number of periods is indicated by the keyword **nper** or **numPeriods** followed by an integer for the number of periods.

4.7.5 Group covariate definitions

If the data contain group covariates, they are defined after the number of periods definition. There is one definition for each group covariate, as follows:

1. The keyword **gcov** or **group_covariate.**
2. The name, or label, for the group covariate.
3. Optionally, one of the following keywords
    - **time_variant** indicates that the group covariate has a different value in each period.
    - **time_invariant** indicates that the group covariate has the same value across all periods.

    If neither keyword is specified, the covariate is assumed to be time invariant.
4. The values of the covariate. If time variant, there must be one value for each population and each period; if time invariant, there must be one value for each population.

4.7.6 Individual covariate definitions

If the data contain individual covariates, there must be one definition for each individual covariate, as follows:

1. The keyword icov or individual_covariate.
2. The name, or label, for the individual covariate.

The value for each individual covariate for each individual is specified in the capture history line.

4.7.7 Population names

One label per population to be used in reports. If omitted, the names default to numbers 1 through n (n = number of populations).

4.7.8 Number tagged

The number tagged is specified by the keyword **ntag** or **number_tagged**, followed by a number for each population.

4.7.9 Tag ID flag

The keyword **tagID** followed by the keyword **absent** indicates that there are no tag IDs in the data and each capture history line begins with the capture history; if the **tagID** keyword is followed by the keyword **present**, the first field of the capture history line is the tag ID. If the **tagID** keyword is not used, the tag IDS are assumed to be present.

4.7.10 Capture history line

The capture history lines are indicated with the keyword capthist or captureHistories, followed by a capture history line for each individual in each population. The individuals must be grouped by populations. For example, if there are two populations specified as

```
number Tagged
250 300
```

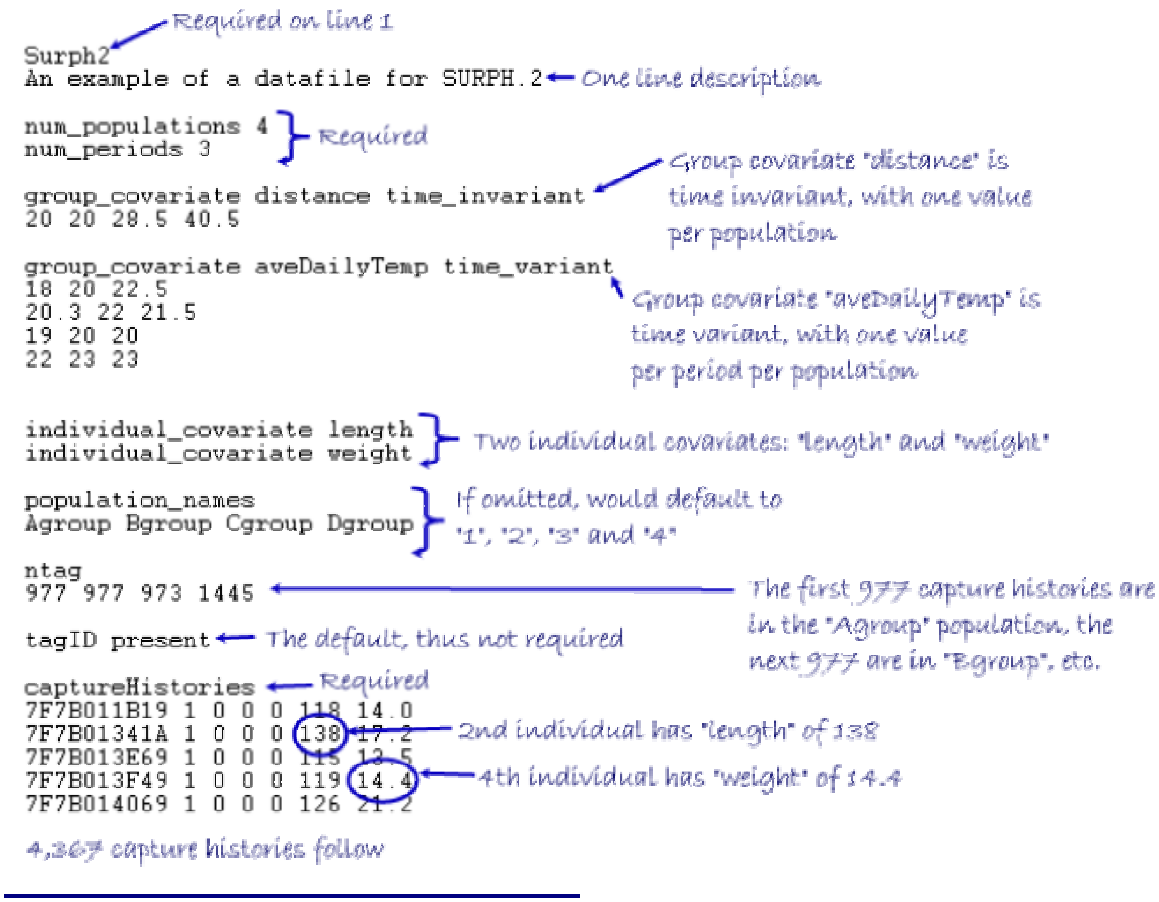
then there must be 550 capture history lines, with the first 250 belonging to the first population, and the remaining 300 belonging to the second population.

Each capture history line is structured as follows:

1. The tag ID if the tagID flag is set to present. Otherwise, the capture history line begins with the capture history, below.
2. The capture history.
3. The individual covariate values. There must be one value for each individual covariate, and they must be specified in the order they are defined.

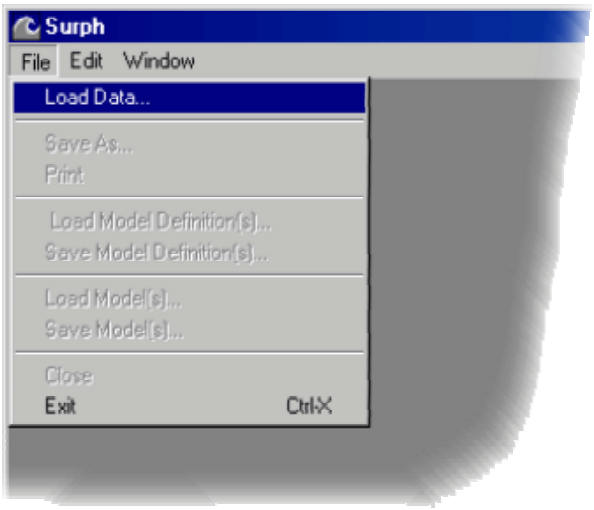
4.7.11 Example: SURPH input data file

Below is an example of a SURPH input data file for a study with 4 populations, 3 periods, 3 group covariates, and 2 individual covariates.


Required on line 1
Surph2
An example of a datafile for SURPH.2
One line description
num_populations 4
num_periods 3
Required
group_covariate distance time_invariant
20 20 28.5 40.5
Group covariate "distance" is time invariant, with one value per population
group_covariate aveDailyTemp time_variant
18 20 22.5
20.3 22 21.5
19 20 20
22 23 23
Group covariate "aveDailyTemp" is time variant, with one value per period per population
individual_covariate length
individual_covariate weight
Two individual covariates: "length" and "weight"
population_names
Agroup Bgroup Cgroup Dgroup
If omitted, would default to "1", "2", "3" and "4"
ntag
977 977 973 1445
The first 977 capture histories are in the "Agroup" population, the next 977 are in "Bgroup", etc.
tagID present
The default, thus not required
captureHistories
Required
7F7B011B19 1 0 0 0 118 14.0
7F7B01341A 1 0 0 0 138 17.2
7F7B013E69 1 0 0 0 115 13.5
7F7B013F49 1 0 0 0 119 14.4
7F7B014069 1 0 0 0 126 21.2
2nd individual has "length" of 138
4th individual has "weight" of 14.4
4,367 capture histories follow

4.8 File Operations

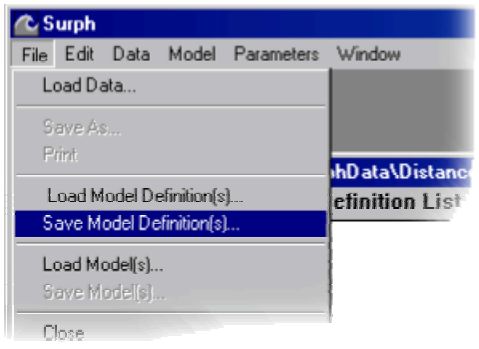
4.8.1 SURPH 2.1 - Load input data

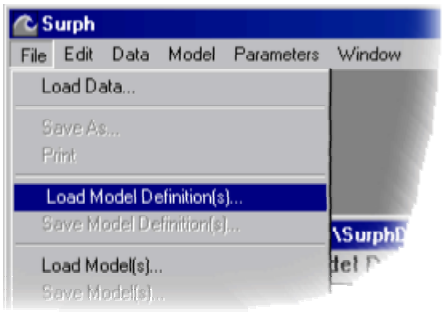


Once the data have been successfully loaded, the SURPH Dialog Window for the selected data type will appear.

Note: The input data file format is different for SURPH 2.1 than for previous versions. A conversion utility is provided to convert older input data files to the new format.

4.8.2 Save and load model definitions

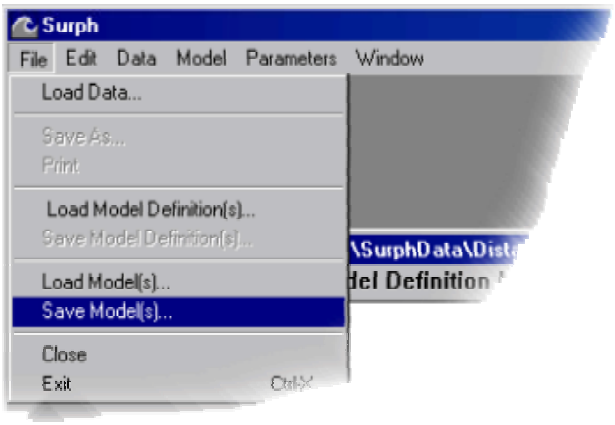


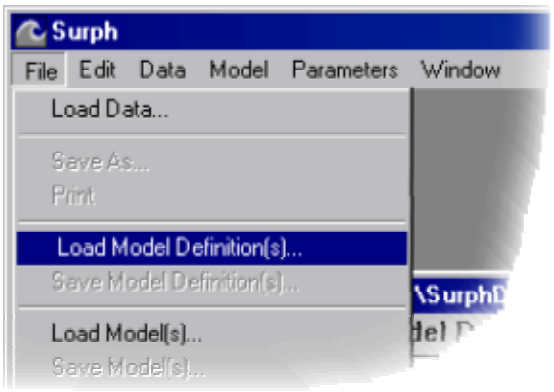


Model Definitions may be saved to a file and loaded at a later SURPH 2.1 session. They are stored as text files that may be edited by the user with a standard text editor (see

Model Definition format). This may be desirable in cases where there are so many populations and/or periods that displaying the model setup on a monitor would be difficult.

4.8.3 Save and load models





Once a Model has been created with SURPH 2.1, it may be saved to a file and then loaded at a later SURPH 2.1 session.

5.0 References

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In B. N. Prtrov and F. Csaki [ed.] 2nd International Symposium on Information Theory. Akademiai Kiado, Budapest.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer, New York. 353 pp.

Burnham, K. P, D. R. Anderson, G. C. White, C. Brownie, and K. H. Pollock. 1987. Design and analysis methods for fish survival experiments based on release-recapture. Am. Fish. Soc. Monograph No. 5, 437 pp.

Cormack, R. M. 1964. Estimates of survival from the sighting of marked animals. Biometrika 51: 429-438.

Jolly, G. M. 1965. Explicit estimates from capture-recapture data with both death and immigrations--stochastic model. Biometrika 52: 225-47.

Manly, B. F. J., and M. J. Parr. 1968. A new method of estimating population size, survivorship and birth rate from capture-recapture data. Trans. Soc. Brit. Ent. 18:81-89.

McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models. 2nd edition. Chapman and Hall, London. 511 pp.

Sakamoto, Y., M. Ishiguro, and G. Kitagawa. 1987. Akaike information criterion statistics. KTK Scientific Publishers, Tokyo. 290 pp. (Transl. from Japanese by Johoryo Tokeigaku.)

Seber, G. A. F. 1965. A note on the multiple recapture census. Biometrika 52: 249-52.

Smith, S. G. 1991. Assessing hazards in wild populations using auxiliary variables in tag-release models. Ph.D. dissertation, University of Washington, Seattle, WA. 351 pp.

White, G. C., and R. A. Garrott. 1990. Analysis of wildlife radio-tracking data. Academic Press, San Diego, CA. 383 pp.